Published in IET Control Theory and Applications
Received on 22nd August 2012
Revised on 10th April 2013
Accepted on 22nd May 2013
doi: 10.1049/iet-cta.2012.0653


ISSN 1751-8644

# Closed-loop subspace identification methods: an overview

*Gijs van der Veen*[1], *Jan-Willem van Wingerden*[1], *Marco Bergamasco*[2], *Marco Lovera*[2], *Michel Verhaegen*[1]

[1] *Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands*
[2] *Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci, 20133 Milano, Italy*
*E-mail: g.j.vanderveen@tudelft.nl*

**Abstract:** In this study, the authors present an overview of closed-loop subspace identification methods found in the recent literature. Since a significant number of algorithms has appeared over the last decade, the authors highlight some of the key algorithms that can be shown to have a common origin in autoregressive modelling. Many of the algorithms found in the literature are variants on the algorithms that are discussed here. In this study, the aim is to give a clear overview of some of the more successful methods presented throughout the last decade. Furthermore, the authors retrace these methods to a common origin and show how they differ. The methods are compared both on the basis of simulation examples and real data. Although the main focus in the literature has been on the identification of discrete-time models, identification of continuous-time models is also of practical interest. Hence, the authors also provide an overview of the continuous-time formulation of the identification framework.

## 1 Introduction

Closed-loop subspace identification of linear systems is of great practical interest for a number of reasons. Linear models are often required for (model-based) control design and, by directly using measured data, system identification overcomes some of the limitations of first principles modelling. Often, simplifying assumptions are made and limited knowledge of true physical parameters is available. For these reasons, system identification may provide more accurate estimates of natural frequencies and input–output gains and hence may also complement modelling on the basis of first principles. In particular, the advantage of subspace methods compared with prediction-error methods [1] has long been recognised in the context of multivariable systems. For these systems, the parametrisation of a prediction-error model structure often leads to an error criterion that is not convex in the parameters. In contrast, the successful closed-loop subspace identification methods developed in recent years consist of a sequence of linear least-squares problems and a model reduction step. In fact, these methods combine prediction-error identification – the estimation of a high-order auto regressive with external input (ARX) model – as an initial step with typical subspace-related subsequent steps [2–5]. Furthermore, regarding the closed-loop nature of such methods, one may observe that in many practical cases feedback is indeed present. On the one hand, this may be necessary owing to instability of the open-loop plant, tight process tolerances, limited access to internal signals of the system or the requirement to stay close to an equilibrium around which one can consider the behaviour of the plant linear. On the other hand, in the literature on identification for control [6] it has often been pointed out that is is desirable to identify a system under circumstances that are close to the real application – that is, in closed-loop – since this results in improved estimation of the dynamics, in particular around the cross-over frequency.

In a closed-loop setting the input signal to the system is typically correlated with the process and measurement noise sources. The presence of correlation because of feedback of stochastic signals (e.g. the feedback of process and/or measurement noise) has traditionally hampered subspace identification in achieving consistent estimates. First efforts to develop subspace methods for data obtained under closed-loop conditions were made in the mid-1990s, soon after the development of the main open-loop subspace methods, see for instance [7–10]. The method developed in [7] shows similarities to the 'joint input–output method' well-known in closed-loop prediction-error identification [11–13], by combining identification of the closed-loop system with knowledge about the controller. In [8], closed-loop subspace identification of a restricted class of closed-loop systems is considered by means of instrumental variables (IV). In [9] the N4SID class of open-loop methods is extended to closed-loop systems, but certain knowledge about the controller is required.

One could state that the ideas presented in [10] have been pivotal to reaching the current state of the art. In that article, high-order ARX modelling was first proposed as a means to deal with correlation issues because of operation

in closed-loop, and it is now a feature of the state-of-the-art algorithms. In the year 2003, several articles appeared that again considered the problem of identification in feedback [3, 14–16]. The article by Qin and Ljung [3] described the first 'innovation estimation' algorithm, in which the first step is to estimate the innovations process. In the article by Jansson [14] the construction of a state predictor as it is currently used in the closed-loop state estimation algorithms, such as predictor-based subspace identification (PBSID), was first considered. The articles by Chiuso and Picci in that same year [15, 16] provided much of the theoretical insight behind these methods, discussing how to deal with feedback models.

While, as we mentioned, Jansson [14] already considered the construction of a state predictor, it was not until 2007 [2] that these developments were combined with the estimation of an ARX model as was proposed earlier in [10]. This resulted in the efficient PBSID algorithm [2], which is currently one of the most promising solutions in closed-loop subspace identification. The innovation estimation algorithm by Qin and Ljung has also seen several further developments over the last decade, resulting in simpler and more efficient implementations [4, 5, 17].

It is interesting to note that all these methods have in common that they rely on estimating a high-order ARX structure to start with. As pointed out by Chiuso [18], the developments in these two broad classes of subspace methods (the state estimation and innovation estimation algorithms) can be seen as a significant step forward towards a satisfactory solution for closed-loop subspace identification problems.

Several other modifications to the existing subspace algorithms, to deal with the closed-loop identification problem, have been proposed in parallel, see for instance [19–21]. In [19] a method is presented that is analogous to the indirect 'two-stage method' in prediction-error identification [13]. In [20] a joint input–output method is presented similar to [7], which focuses on the deterministic subsystems. Finally, in [21] an IV approach was developed, which requires an estimate of the noise model that is not available a priori (see [22] for a recent overview of these IV methods for prediction-error identification.). Hence, an iterative procedure is proposed to estimate the model.

In prediction-error identification the available approaches can typically be classified as 'direct', 'indirect' and 'joint input–output' approaches [13]. In this paper, we focus on closed-loop susbpace methods that we would like to classify as direct methods [2, 4, 5, 17]. Although these methods consist of several steps, they operate directly on the available input–output data. These 'direct' subspace methods have the advantage that they place the fewest restrictions on the feedback mechanism. The main drawbacks of the indirect and joint input–output methods [7, 9, 19–21] are that linearity of the closed-loop system (not just the plant) must be assumed and that care must be taken with pole-zero cancellations between the plant and the controller.

Based on the foregoing discussion it can be concluded that the field of closed-loop subspace identification has been very active. So active in fact, that a first-time survey of algorithms and methodologies may be daunting to readers who are new to the field. The aim of this paper is to give an overview of some of the more successful methods presented throughout the last decade. Furthermore, we retrace these methods to a common origin and show how they differ. A natural question to be asked is which method is the best for a particular purpose. This is a hard question, since indeed it depends on the purpose of the identified model [6], the type of the underlying system [23] – particularly using finite sequences – and experimental conditions. Many authors have addressed this issue from a theoretical perspective, regarding asymptotic consistency and variance [23–26]. When dealing with practical conditions, that is, finite-length sequences and systems of unknown order (or distributed-parameter systems), the various methods discussed here may perform rather differently.

Although most of the research in system identification focuses on discrete-time models, in many situations of practical interest (such as, e.g. aircraft and rotorcraft identification, see e.g. [27–29]) the direct estimation of the parameters of a continuous-time model from sampled input–output data is desirable, and for that dedicated methods and tools are needed. In addition, there exist special cases in which identifying discrete-time models can be critical, such as the identification of stiff systems or the use of non-equidistantly sampled data, which make it necessary to develop special algorithms that can deal with these cases. The development of identification methods for continuous-time models is a challenge of its own, and has been studied extensively (see, e.g. the recent book [30] and references therein, and the recent special issue [31]).

The problem of closed-loop subspace identification in continuous-time has been first considered in the literature in [32], where the application of the errors-in-variables approach of [8] is proposed to deal with correlation in a continuous-time setting.

More recently, see [33–35], novel continuous-time subspace model identification (SMI) schemes, based on the derivation of PBSID-like algorithms within the all-pass domains proposed in [36, 37] and relying, respectively, on Laguerre filtering and Laguerre projections of the sampled input–output data have been proposed.

The paper proceeds as follows: first, in Section 2, we present the predictor framework that is common to the methods discussed in this paper. In Section 3, it is shown how the results from Section 2 can be used in several ways to arrive at a state-space realisation and which user choices must be made. This leads us to the essential features of many of the different methods considered in the literature. Section 4 is devoted to the discussion of closed-loop subspace identification for continuous-time models. Finally, in Section 5 we describe the results of some experimental studies performed using the presented algorithms. In the first two examples we consider two numeric examples to study the differences between the basic algorithms. In the third example, we consider closed-loop identification of beam dynamics and we consider the quality of the identified models on the basis of Monte Carlo experiments. In the fourth example, continuous-time models are identified of the same system. The paper concludes with a brief discussion of the results.

## 2 Discrete-time identification framework

In this section, we present the framework for closed-loop identification of discrete-time systems. It is assumed throughout that the system to be identified is a finite-dimensional, linear, time-invariant system, subject to measurement and/or process noise. Based on these assumptions, the system admits a discrete-time innovation state-space representation $\mathcal{P}_{\mathrm{d}}$ given by

$$\mathcal{P}_{\mathrm{d}}: \begin{cases} \boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{K}\boldsymbol{e}_k & \text{(1a)} \\ \boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{D}\boldsymbol{u}_k + \boldsymbol{e}_k & \text{(1b)} \end{cases}$$



*IET Control Theory Appl.*, 2013, Vol. 7, Iss. 10, pp. 1339–1358
doi: 10.1049/iet-cta.2012.0653

with $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_u}$, $\boldsymbol{K} \in \mathbb{R}^{n\times n_y}$, $\boldsymbol{C} \in \mathbb{R}^{n_y\times n}$ and $\boldsymbol{D} \in \mathbb{R}^{n_y\times n_u}$. The vectors $\boldsymbol{x}_k \in \mathbb{R}^n$, $\boldsymbol{u}_k \in \mathbb{R}^{n_u}$, $\boldsymbol{y}_k \in \mathbb{R}^{n_y}$ and $\boldsymbol{e}_k \in \mathbb{R}^{n_y}$ are the state vector, input signal, output signal and innovation signal, respectively. The matrix $\boldsymbol{K}$ is the Kalman gain. The innovation sequence $\boldsymbol{e}_k$ is an ergodic zero-mean white noise sequence with covariance matrix $\mathbb{E}\{\boldsymbol{e}_j\boldsymbol{e}_k^\top\} = \boldsymbol{W}\delta_{jk}$, with $\boldsymbol{W} \succ 0$.

It is assumed that the pair $(\boldsymbol{A}, \boldsymbol{C})$ is observable and the pair $(\boldsymbol{A}, [\boldsymbol{B}\ \boldsymbol{K}\boldsymbol{W}^{\frac{1}{2}}])$ is reachable. While $\boldsymbol{A} - \boldsymbol{K}\boldsymbol{C}$ could have eigenvalues on the unit circle, in this context we assume that all eigenvalues are inside the unit circle, to allow consistent estimation of the observer Markov parameters (see Section 2.2). The state-space model (1) is the innovation model associated with a generic state-space model.

In representation (1), $\boldsymbol{e}_k$ may be eliminated from the first equation to yield a system description in one-step-ahead predictor form

$$\boldsymbol{x}_{k+1} = \tilde{\boldsymbol{A}}\boldsymbol{x}_k + \tilde{\boldsymbol{B}}\boldsymbol{u}_k + \boldsymbol{K}\boldsymbol{y}_k \tag{2a}$$

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{D}\boldsymbol{u}_k + \boldsymbol{e}_k \tag{2b}$$

where $\tilde{\boldsymbol{A}} \equiv \boldsymbol{A} - \boldsymbol{K}\boldsymbol{C}$ and $\tilde{\boldsymbol{B}} \equiv \boldsymbol{B} - \boldsymbol{K}\boldsymbol{D}$ have been introduced for brevity. We will use the notation $\tilde{\cdot}$ whenever a parameter refers to the predictor model (2). This representation forms the basis for the PBSID framework [38].

The goal of the following subsections will be to deliver a unified presentation of the closed-loop subspace identification methods, showing that they have a common origin. Our aim is to provide the reader with a good understanding of key steps in the algorithms, both for implementation and analysis purposes.

For reference purposes, we define the identification problem below in Problem 1. It is assumed that the plant $\mathcal{P}$ operates in closed-loop with a, not necessarily linear, controller $\mathcal{C}$ as shown in Fig. 1. In this figure, it is already assumed that the noise effects are modelled as if originating from a filtered innovation sequence. For convenience, we define a combined reference signal (without assuming that the controller is linear time invariant (LTI))

$$\boldsymbol{r}_k = \boldsymbol{r}_{1,k} + \mathcal{C}(\boldsymbol{r}_{2,k})$$

Note that recent research has demonstrated that care must be taken when identifying an LTI system controlled by a non-linear feedback mechanism. In [39] it is demonstrated that problems may occur if the controller is non-linear and if the true noise model is not stably invertible, for example, if it is non-minimum phase.

We assume that the feedback system is well-posed, implying that the output is uniquely determined by the states. The feedback system is well-posed if the controller or the plant or both have no direct feedthrough component. If the system and the controller are both LTI, the condition for well-posedness is satisfied if $\boldsymbol{I}_{n_y} + \boldsymbol{D}\boldsymbol{D}_c$ is non-singular, where $\boldsymbol{D}_c$ is the feedthrough matrix of the controller [9, 40]. Hence, without any means of constraining the structure of $\boldsymbol{D}$, we can either choose to include it or not in the identification procedure, depending on the feedback system having direct feedthrough or not. Note that a correct choice is necessary to obtain consistent estimates of the Markov parameters; see Section 2.2.3.

It is further assumed that the reference signal $\boldsymbol{r}_k$ is such that $\boldsymbol{u}_k$ and $\boldsymbol{y}_k$ are jointly persistently exciting of sufficiently high order (see Section 2.2.3 for more details).

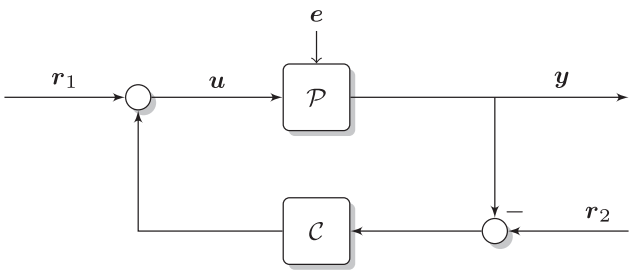


**Fig. 1** *Closed-loop configuration* $\Sigma$ *considered in Problem 1*

*Problem 2.1:* **Discrete-time subspace identification problem** Based on a finite set of input and output data $\{\boldsymbol{u}_k, \boldsymbol{y}_k\}_{k=0}^{N-1}$ obtained from a system $\Sigma$, estimate the order $n$ of the discrete-time system $\mathcal{P}_d$ and the associated system matrices $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{K})$ up to a similarity transformation.

### 2.1 Preliminaries and notation

Before deriving the data equations for subspace identification, we will introduce some notation. We introduce a stacked sample of input and output data $\boldsymbol{z}_k$ according to

$$\boldsymbol{z}_k = \begin{bmatrix} \boldsymbol{u}_k \\ \boldsymbol{y}_k \end{bmatrix}$$

The stacked vector $\boldsymbol{z}_k^{(p)}$ is defined as

$$\boldsymbol{z}_k^{(p)} = \begin{bmatrix} \boldsymbol{z}_{k-p}^\top, & \boldsymbol{z}_{k-p+1}^\top, & \ldots, & \boldsymbol{z}_{k-1}^\top \end{bmatrix}^\top$$

where $p$ denotes the 'past window' size. We also define a reversed extended controllability matrix $\tilde{\mathcal{K}}^{(p)}$

$$\tilde{\mathcal{K}}^{(p)} = \begin{bmatrix} \tilde{\boldsymbol{A}}^{p-1}\bar{\boldsymbol{B}}, & \tilde{\boldsymbol{A}}^{p-2}\bar{\boldsymbol{B}}, & \ldots, & \bar{\boldsymbol{B}} \end{bmatrix} \tag{3}$$

where we have defined $\bar{\boldsymbol{B}} = [\tilde{\boldsymbol{B}}, \boldsymbol{K}]$ for brevity. We emphasise that this matrix contains parameters pertaining to the 'predictor' representation (2). We will further denote block Hankel matrices constructed from data sequences according to

$$\boldsymbol{Y}_{i,s,N} = \begin{bmatrix} \boldsymbol{y}_i & \boldsymbol{y}_{i+1} & \cdots & \boldsymbol{y}_{i+N-1} \\ \boldsymbol{y}_{i+1} & \boldsymbol{y}_{i+2} & \cdots & \boldsymbol{y}_{i+N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{y}_{i+s-1} & \boldsymbol{y}_{i+s} & \cdots & \boldsymbol{y}_{i+N+s-2} \end{bmatrix}$$

such that $\boldsymbol{Y}_{i,s,N}$ has $\boldsymbol{y}_i$ as its first element and possesses $s$ block rows and $N$ columns. We will sometimes consider block-row matrices, that is, with $s = 1$, which we shall denote by $\boldsymbol{Y}_{i,N}$. Finally, we define the block-Toeplitz matrix $\boldsymbol{H}^{(f)}(\boldsymbol{B}, \boldsymbol{D})$ pertaining to the innovation (1) model, which is to be used later when the Multivariable Output-Error State-sPace (MOESP) algorithm is outlined

$$\boldsymbol{H}^{(f)}(\boldsymbol{B}, \boldsymbol{D}) = \begin{bmatrix} \boldsymbol{D} & 0 & 0 & \cdots & 0 \\ \boldsymbol{C}\boldsymbol{B} & \boldsymbol{D} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ \boldsymbol{C}\boldsymbol{A}^{f-2}\boldsymbol{B} & \boldsymbol{C}\boldsymbol{A}^{f-3}\boldsymbol{B} & \cdots & \boldsymbol{C}\boldsymbol{B} & \boldsymbol{D} \end{bmatrix}$$

Likewise, we define the matrix $\boldsymbol{H}^{(f)}(\boldsymbol{K}, 0)$.

### 2.2 Data equations

In this section, we derive the data equation that is common to many of the closed-loop subspace algorithms. Starting from some initial state $\boldsymbol{x}_k$, the state equation (2a) can be propagated forward in time, resulting in the expression

$$\boldsymbol{x}_{k+p} = \tilde{\boldsymbol{A}}^p \boldsymbol{x}_k + \tilde{\mathcal{K}}^{(p)} \boldsymbol{z}_{k+p}^{(p)}. \tag{4}$$

Based on (4) and the output equation (2b), the output at time $k+p$ can then be written as

$$\boldsymbol{y}_{k+p} = \boldsymbol{C}\tilde{\boldsymbol{A}}^p \boldsymbol{x}_k + \boldsymbol{C}\tilde{\mathcal{K}}^{(p)} \boldsymbol{z}_{k+p}^{(p)} + \boldsymbol{D}\boldsymbol{u}_{k+p} + \boldsymbol{e}_{k+p} \tag{5}$$

By the assumption that $\tilde{\boldsymbol{A}}$ has all its eigenvalues inside the open unit disc, the term $\tilde{\boldsymbol{A}}^p$ can be made arbitrarily small, that is, $\|\tilde{\boldsymbol{A}}^p\|_2 \simeq 0$, by choosing $p$ sufficiently large. [See also Section 3.5.1 further on regarding this issue in relation to finite data lengths.] For that reason, the first term on the right-hand sides of (4) and (5) will be neglected. Since all further algorithms are based on this assumption, we introduce it formally.

*Assumption 2.1 Negligible bias:* It is assumed that the choice of the past window size $p$ results in $\tilde{\boldsymbol{A}}^p = 0$.

Depending on the number of samples available and based on Assumption 2.1, (5) can be repeated to obtain expressions for $\boldsymbol{y}_p$ up to $\boldsymbol{y}_{N-1}$, resulting in

$$\boldsymbol{Y}_{p,N_p} = \boldsymbol{C}\tilde{\mathcal{K}}^{(p)} \boldsymbol{Z}_{0,p,N_p} + \boldsymbol{D}\boldsymbol{U}_{p,N_p} + \boldsymbol{E}_{p,N_p} \tag{6}$$

Here, we have defined $N_p = N - p$ for brevity. In the remainder of this article, the equality in (6) is understood to hold under Assumption 2.1. As noted before, the feedthrough term $\boldsymbol{D}$ should only be included when the feedback loop contains at least a one-sample delay (i.e. has no direct feedthrough) to retain consistency of the identification problem. From (6) it is clear that if the controller has direct feedthrough, $\boldsymbol{U}_{p,N_p}$ is correlated with $\boldsymbol{E}_{p,N_p}$ and the Markov parameters can no longer be estimated consistently.

*Remark 2.1:* If one leaves out the input terms in the assumed model structure (1), it is possible to identify a stochastic model (spectral factor) of a process driven by white noise on the basis of output measurements only. In these cases, however, there is no need to apply a closed-loop identification method. See, for example, [41, Chapter 3], [40, 42].

#### 2.2.1 Relation to the ARX model structure:

Taking a closer look at the data equation (5), neglecting the first term on the right-hand side (for $p \to \infty$), it is seen to have a vector-ARX (VARX) structure. Usually, an ARX model structure prescribes a severely restrictive noise model because it forces the system and noise model to have a common set of poles as seen from the following equation

$$\boldsymbol{A}(z)\boldsymbol{y}_k = \boldsymbol{B}(z)\boldsymbol{u}_k + \boldsymbol{e}_k \tag{7}$$

with $z^{-1}$ the unit backshift operator and

$$\boldsymbol{A}(z) = \boldsymbol{I} - \boldsymbol{a}_1 z^{-1} - \cdots - \boldsymbol{a}_p z^{-p}$$
$$\boldsymbol{B}(z) = \boldsymbol{b}_0 + \boldsymbol{b}_1 z^{-1} + \cdots + \boldsymbol{b}_p z^{-p}$$

In this context, based on the assumption that $p$ is chosen sufficiently large and working with the predictor form (2), it follows that the high-order VARX model is fully equivalent to the predictor model. The fact that a high-order ARX model can approximate a predictor model with arbitrary accuracy is well-known in prediction-error identification; cf. [43, example 10.11].

Regarding the ARX model structure defined in (7), the parameters $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ can explicitly be given as the Markov parameters of the predictor form (2)

$$\boldsymbol{a}_i = \boldsymbol{C}\tilde{\boldsymbol{A}}^{i-1}\boldsymbol{K}, \quad \text{for } i = 1 \ldots p \tag{8a}$$
$$\boldsymbol{b}_i = \boldsymbol{C}\tilde{\boldsymbol{A}}^{i-1}\tilde{\boldsymbol{B}}, \quad \text{for } i = 1 \ldots p \tag{8b}$$
$$\boldsymbol{b}_0 = \boldsymbol{D} \tag{8c}$$

This follows by a direct comparison of (7) with (5) after neglecting the contribution of the initial state (i.e. for $p \to \infty$).

#### 2.2.2 Closed-loop identification issues:

Traditional formulations of subspace identification methods often require the plant to operate under open-loop conditions. If such methods are applied to data obtained under closed-loop conditions, the fact that the input signal to the system is correlated with the noise processes is disregarded or neglected. The implicit assumption in such methods is that the input signal $\boldsymbol{u}_k$ is uncorrelated with the past noise process $\boldsymbol{e}_k$. In a closed-loop situation, however, it is clearly seen that this condition is violated

$$\mathbb{E}\{\boldsymbol{u}_k \boldsymbol{e}_j^\top\} \neq 0 \quad \text{for } j < k$$

Over the last two decades, several strategies have been introduced to deal with this issue, of which we mention a few:

1. Use an open-loop subspace identification method and either accept the bias on the system estimate or use it as an initial model in a prediction-error method [1]. Many prediction-error methods are available to deal with closed-loop situations [13, 44]. It is nevertheless of interest to use a subspace method that is better suited to closed-loop data to obtain a better initial estimate for prediction-error methods, which rely on solving a non-convex optimisation problem;
2. Use a particularly chosen reference signal $\boldsymbol{r}_k$ as, for example, discussed in [8] to retain certain consistency properties of the identification algorithm. In [45], it is argued that, if the feedback is still an open choice, correlation issues because of feedback can also be remedied by a particular choice of the feedback mechanism, for example, if state feedback based on a Kalman filter is applied. The states estimated by a Kalman filter are (in an ideal setting) uncorrelated with the innovations. Hence, the feedback signal contains no feedback of the noise process(es). Obviously, this approach only lends itself to certain design cases. On page 9 and onwards in [46], the effect of specific input signals on the properties of the closed-loop identification problem is also discussed.
3. Use knowledge of the controller, which is then often assumed to be LTI, to achieve consistent estimates [7, 9, 20].
4. Modify the subspace identification algorithms so as to achieve identification methods that are directly suited to closed-loop data [2–5, 10, 14, 17, 19, 21, 47]. In this paper, we treat the most dominant developments in closed-loop subspace identification, given by the methods that perform high-order ARX modelling followed by a second step that




doi: 10.1049/iet-cta.2012.0653

includes model reduction [2–5, 10, 14, 17]. In prediction-error identification, the available approaches can typically be classified as 'direct', 'indirect' and 'joint input–output' approaches [13]. In this paper, we focus on closed-loop subspace methods that we would like to classify as direct methods [2–5, 10, 14, 17]. Although these methods consist of several steps, they operate directly on the available input–output data and have the advantage that they place the fewest restrictions on the feedback mechanism.

In the PBSID framework, resulting in (VARX) data equations of the form (6), the estimation is not affected by correlation issues, by segregating the data into collections of 'past' and 'future' samples. Thus, asymptotically in the number of samples $N$ and the past window size $p$, the parameters can be consistently estimated.

*2.2.3 Estimating the predictor Markov parameters:* Based on the assumption that $\boldsymbol{e}_k$ is the zero-mean white noise innovation sequence and on Assumption 2.1, the predictor Markov parameters in (6) can be consistently estimated in a least-squares sense

$$\min_{[\boldsymbol{C}\tilde{\mathcal{K}}^{(p)}\ \boldsymbol{D}]} \left\| \boldsymbol{Y}_{p,N_p} - \left[\boldsymbol{C}\tilde{\mathcal{K}}^{(p)}\ \boldsymbol{D}\right] \begin{bmatrix} \boldsymbol{Z}_{0,p,N_p} \\ \boldsymbol{U}_{p,N_p} \end{bmatrix} \right\|_F^2 \tag{9}$$

For a full-rank data matrix $\left[\boldsymbol{z}_{0,p,N_p}^{\top},\ \boldsymbol{U}_{p,N_p}^{\top}\right]^{\top}$, the least-squares solution can be found from an RQ decomposition [48] of the data. Performing an RQ factorisation of the stacked data matrices one obtains

$$\begin{bmatrix} \begin{bmatrix} \boldsymbol{Z}_{0,p,N_p} \\ \boldsymbol{U}_{p,N_p} \end{bmatrix} \\ \boldsymbol{Y}_{p,N_p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_{11} & 0 \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_1 \\ \boldsymbol{Q}_2 \end{bmatrix}$$

from which it can be derived that the parameters can be found by solving

$$\boldsymbol{R}_{21} = \widehat{\left[\boldsymbol{C}\tilde{\mathcal{K}}^{(p)}\ \boldsymbol{D}\right]}\boldsymbol{R}_{11}$$

for example, using back-substitution. In addition, using the orthogonality of the rows of $\boldsymbol{Q}$, we may obtain an estimate of $\boldsymbol{E}_{p,N_p}$, according to

$$\hat{\boldsymbol{E}}_{p,N_p} = \boldsymbol{R}_{22}\boldsymbol{Q}_2 \tag{10}$$

This matrix contains an estimate of the innovation sequence of the innovation state-space model (1). Note that the innovation sequence is obtained without explicitly solving the least-squares problem. Note that if $\boldsymbol{D}$ is included, it has now been estimated and thus its estimation will not be considered further on.

Based on the least-squares solution, we now have estimates of the predictor Markov parameters (8) and the innovation sequence $\{\boldsymbol{e}_k\}_{k=p}^{N-1}$.

It is obvious that uniqueness of the parameter estimate requires $\bar{\boldsymbol{Z}} = \left[\boldsymbol{z}_{0,p,N_p}^{\top},\ \boldsymbol{U}_{p,N_p}^{\top}\right]^{\top}$ to be of full rank. The information matrix $\mathcal{I} = \bar{\boldsymbol{Z}}\bar{\boldsymbol{Z}}^{\top}$ related to the least-squares solution will then be positive definite. This requirement in turn depends on the experimental data, and thus on the true system, reference excitation and the nature of the feedback mechanism [49]. For interesting accounts regarding model identifiability and experiment requirements, see, for example, [49, 50]. We emphasise that in practice, purely from an identification point of view it is usually advantageous to make the reference perturbations as large as possible within the limitations of the system and the requirement to stay close to an operating point around which the system behaves linearly. In many cases, it may be necessary to be more judicious in the choice of perturbations: this leads to the topic of least costly identification [6, 51].

*2.2.4 Least-squares sensitivity:* It is interesting to note that using the estimate $\widehat{\boldsymbol{E}}_{p,N_p}$ instead of the Markov parameters may be advantageous in certain cases where the data matrix $\bar{\boldsymbol{Z}} = \left[\boldsymbol{z}_{0,p,N_p}^{\top},\ \boldsymbol{U}_{p,N_p}^{\top}\right]^{\top}$ is poorly conditioned. This may occur when the joint input–output data is not rich enough, for example, owing to the nature of the experiment, poor excitation of reference signals or a low-complexity feedback path [49]. In fact, it can be shown that the estimate of the least-squares residual of (9) is less sensitive to ill-conditioning than the estimate of the parameters themselves. The worst-case sensitivities of the two estimates are related to the condition number $\varrho$ of the data matrix as follows [48]

$$\text{parameters: } \frac{\left\| \Delta\widehat{\left[\boldsymbol{C}\tilde{\mathcal{K}}^{(p)},\ \boldsymbol{D}\right]} \right\|_2}{\left\| \left[\boldsymbol{C}\tilde{\mathcal{K}}^{(p)},\ \boldsymbol{D}\right] \right\|_2} \propto \varrho\left(\bar{\boldsymbol{Z}}\right)^2$$

$$\text{residual: } \frac{\left\| \Delta\widehat{\boldsymbol{E}}_{p,N_p} \right\|_2}{\left\| \boldsymbol{Y}_{p,N_p} \right\|_2} \propto \varrho\left(\bar{\boldsymbol{Z}}\right)$$

where $\varrho(\cdot)$ denotes the condition number. This shows that the estimate of the parameters may be far more sensitive to ill-conditioning than the estimate of the residual.

It is hard to draw further general conclusions from these facts, in particular since one of the estimates is used in state reconstruction algorithms (e.g. $\text{PBSID}_{\text{opt}}$, Section 3.3), whereas the other is used in algorithms, which estimate the observability matrix (e.g. CLMOESP, Section 3.4). It is expected, however, that in cases where the data matrix is severely ill-conditioned, the estimate of the innovation sequence may be more reliable than that of the Markov parameters, in particular for the subsequent step of estimating the observability matrix and hence the eigenvalues. Section 5.2 discusses an example that illustrates this possible effect.

## 3 Obtaining a state-space realisation

In the previous section, estimates were obtained for the Markov parameters and the innovation signal pertaining to the predictor model (2). In the following subsections, we consider four different methods to arrive at a solution to the identification problem (Problem 1) based on these estimates. Fig. 2 schematically depicts the different routes from input–output data to an identified model with references to the appropriate subsections.

### 3.1 Direct parametrisation

It is possible to directly obtain a non-minimal state-space model, by casting the estimated ARX model parameters into

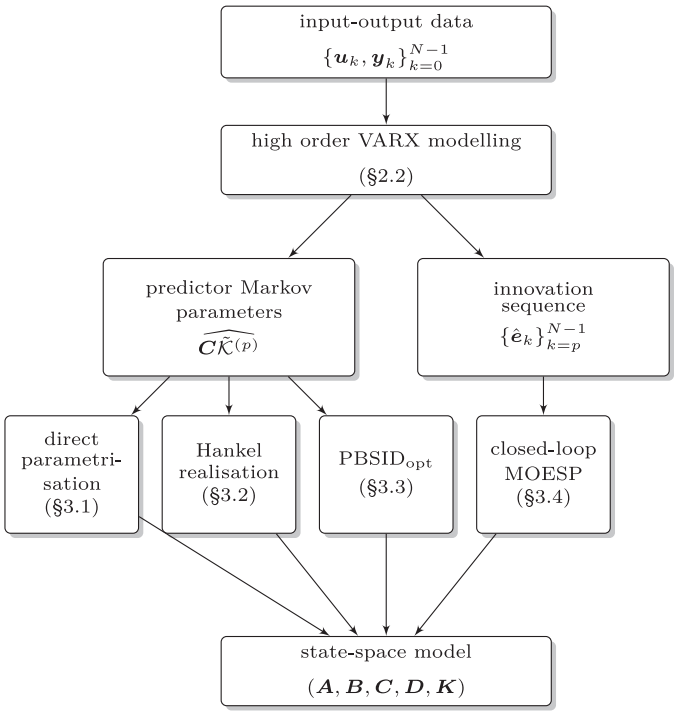


**Fig. 2** *Schematic representation of the relation between the different closed-loop subspace algorithms*

a state-space parametrisation of order $(n_y + n_u)p$ (11)

$$z_{k+1} = \left[\begin{array}{cccc} 0 & I & & \\ & \ddots & \ddots & \\ & & 0 & I \\ \hline 0 & \cdots & \cdots & 0 \\ \hline & \multicolumn{2}{c}{\widehat{C\tilde{\mathcal{K}}^{(p)}}} & \end{array}\right] z_k + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ I \\ D \end{bmatrix} u_k + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ I \end{bmatrix} e_k \quad (11a)$$

$$y_k = \widehat{C\tilde{\mathcal{K}}^{(p)}} z_k + Du_k + e_k \quad (11b)$$

Since the VARX parameters are directly placed in the state-space matrices, we shall refer to this parametrisation as the 'direct parametrisation'. The order of this model could subsequently be reduced using a model reduction algorithm. A notable advantage of the direct parametrisation is that the state is measurable, since it is given by delayed samples of input/output data. Furthermore, the variance on the elements of the state-space matrices is directly provided by the least-squares estimate (9). These interesting features were exploited in [52] for purposes of robust state-feedback compensator design. Direct use is also made of the predictor Markov parameters in closed-loop subspace predictive control (SPC) [53]. A drawback of the direct parametrisation is that it is non-minimal and typically has a large state dimension when $p$ is moderate to large. This may be a problem for subsequent control design. Standard model reduction techniques can be applied to reduce the order of this model.

### 3.2 A realisation algorithm

Realisation methods, initiated with the development of the Ho–Kalman realisation algorithm [54–56], are the oldest methods that could be classified under the subspace methods. Whereas earlier approaches depart from a set of Markov parameters or impulse response parameters, later approaches typically start with estimating predictor Markov parameters, or, equivalently, the predictor impulse response, very similar to the VARX step described in Section 2.2.3. In fact, the approach of estimating predictor Markov parameters followed by a realisation step was already presented as early as in 1993 [57]. The approach outlined here is a variation on the Observer/Kalman Filter Identification (OKID) method [58]. In this approach, contrary to the Ho–Kalman approach, first an estimate is obtained of the predictor impulse response, using the fact that this response tends to zero after $p$ steps.

The realisation method relies on forming an 'extended observability-times-controllability' matrix. Let us first introduce the extended observability matrix of the predictor and innovation models

$$\tilde{\Gamma}^{(f)} = \begin{bmatrix} C \\ C\tilde{A} \\ \vdots \\ C\tilde{A}^{f-1} \end{bmatrix}, \quad \Gamma^{(f)} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-1} \end{bmatrix}$$

Given a 'future' window $f > n$, the extended observability-times-controllability matrix $\tilde{\Gamma}^{(f)}\tilde{\mathcal{K}}^{(p)}$ can be constructed, which has the following structure

$$\tilde{\Gamma}^{(f)}\tilde{\mathcal{K}}^{(p)} = \begin{bmatrix} C\tilde{A}^{p-1}\bar{B} & C\tilde{A}^{p-2}\bar{B} & \cdots & C\bar{B} \\ C\tilde{A}^{p}\bar{B} & C\tilde{A}^{p-1}\bar{B} & \cdots & C\tilde{A}\bar{B} \\ \vdots & \vdots & \ddots & \vdots \\ C\tilde{A}^{p+f-2}\bar{B} & C\tilde{A}^{p+f-3}\bar{B} & \cdots & C\tilde{A}^{f-1}\bar{B} \end{bmatrix} \quad (12)$$

Based on the earlier assumption that $\|\tilde{A}^p\|_2 \simeq 0$, the same approximation can be introduced here, resulting in

$$\tilde{\Gamma}^{(f)}\tilde{\mathcal{K}}^{(p)} \approx \begin{bmatrix} C\tilde{A}^{p-1}\bar{B} & C\tilde{A}^{p-2}\bar{B} & \cdots & C\bar{B} \\ 0 & C\tilde{A}^{p-1}\bar{B} & \cdots & C\tilde{A}\bar{B} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & & C\tilde{A}^{f-1}\bar{B} \end{bmatrix} \equiv \begin{bmatrix} \Xi_0 \\ \Xi_1 \\ \vdots \end{bmatrix} \quad (13)$$

Having estimated the predictor Markov parameters $\widehat{C\tilde{\mathcal{K}}^{(p)}}$, it is straightforward to construct this matrix, by noting that each block-row $\Xi_i$ is obtained from the previous by shifting it and padding it with zeroes. The first block-row $\Xi_0$ consists of all predictor Markov parameters, that is, $\Xi_0 \equiv C\tilde{\mathcal{K}}^{(p)}$.

Using the former definition of $\tilde{\mathcal{K}}^{(p)}$ (3), it is possible to derive a recursive expression, which provides the approximate block-rows of the following extended observability-times-controllability matrix

$$\Gamma^{(f)}\tilde{\mathcal{K}}^{(p)} \equiv \begin{bmatrix} \mathcal{H}_0 \\ \mathcal{H}_1 \\ \vdots \\ \mathcal{H}_{f-1} \end{bmatrix} \quad (14)$$

where the recursive expression is given by the authors [53, 57]

$$\mathcal{H}_j = \Xi_j + \sum_{\tau=0}^{j-1} (C\tilde{A}^{j-\tau-1}K)\mathcal{H}_\tau, \quad \mathcal{H}_0 = \Xi_0$$

Based on the assumption of minimality, which holds for both the innovation and predictor representations, it immediately




doi: 10.1049/iet-cta.2012.0653

follows that $\text{rank}(\Gamma^{(f)}\tilde{\mathcal{K}}^{(p)}) = n$. In practice, the number of non-zero singular values is not $n$, because of the fact that we construct the matrix (13) using estimated parameters. Then, an singular value decomposition (SVD) of $\Gamma^{(f)}\tilde{\mathcal{K}}^{(p)}$ can be used to find approximations of $\Gamma^{(f)}$ and $\tilde{\mathcal{K}}^{(p)}$ and the order $n$

$$\Gamma^{(f)}\tilde{\mathcal{K}}^{(p)} \approx \mathcal{U}_n \Sigma_n \mathcal{V}_n^\top$$

so that we may take

$$\Gamma^{(f)} = \mathcal{U}_n, \ \tilde{\mathcal{K}}^{(p)} = \Sigma_n \mathcal{V}_n^\top$$

Estimates of the system matrices can then be obtained as follows: $\boldsymbol{C}$, $\boldsymbol{B}$ and $\boldsymbol{K}$ are simply read off from the appropriate matrices

$$\boldsymbol{C} = \Gamma^{(f)}(1{:}\,n_y, :)$$
$$[\boldsymbol{B} - \boldsymbol{K}\boldsymbol{D}, \ \boldsymbol{K}] = \tilde{\mathcal{K}}^{(p)}(:, (p-1)(n_u + n_y) + 1{:}\,p(n_u + n_y))$$

whereas $\boldsymbol{A}$ is found as the solution to the overdetermined problem

$$\Gamma^{(f)}(1{:}\,(s-1)n_y, :)\boldsymbol{A} = \Gamma^{(f)}(n_y + 1{:}\,sn_y, :) \tag{15}$$

(using MATLAB notation).

### 3.3 Predictor-based subspace identification (PBSID$_{\text{opt}}$)

In predictor-based subspace identification, a predictor for the state sequence is constructed. If an estimate of the state sequence is known, the system matrices can be found directly from two least-squares problems, similar to what is done in the open-loop N4SID class of algorithms, based on the following identities

$$\boldsymbol{X}_{p+1,N_p-1} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} & \boldsymbol{K} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{p,N_p-1} \\ \boldsymbol{U}_{p,N_p-1} \\ \boldsymbol{E}_{p,N_p-1} \end{bmatrix} \tag{16}$$

$$\boldsymbol{Y}_{p,N_p} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{p,N_p} \\ \boldsymbol{U}_{p,N_p} \end{bmatrix} + \boldsymbol{E}_{p,N_p} \tag{17}$$

On the basis of (4) it can be concluded that, neglecting the first term, the product $\tilde{\mathcal{K}}^{(p)}\boldsymbol{Z}_{0,p,N_p}$ represents the state sequence $\boldsymbol{X}_{p,N_p}$. Unfortunately, this product cannot be estimated directly. What is available is an estimate of the parameters $\boldsymbol{C}\tilde{\mathcal{K}}^{(p)}$. As in the OKID algorithm in Section 3.2 we can construct the matrix in (13) using these parameters, now using a 'future' window $f \geq n$ (note that the future window may now be equal to the system order; also see Section 3.5.2). This matrix is used in the PBSID$_{\text{opt}}$ algorithm, whereas the standard PBSID algorithm makes use of the full matrix in (12).

*Remark 3.1:* Following the estimation of the Markov parameters (9) we only have available the Markov parameters required to construct the matrix in (13). It is also possible, however, to construct the full matrix in (12), by solving a sequence of $f$ shifted versions of (9) of increasing order $p, p+1, \ldots$. This is detailed in, for example, [2]. In [2, 38] it was shown that the 'optimised' version described here results in a lower variance than the standard PBSID algorithm.

Thus, having constructed the matrix $\tilde{\Gamma}^{(f)}\tilde{\mathcal{K}}^{(p)}$, the product $\tilde{\Gamma}^{(f)}\tilde{\mathcal{K}}^{(p)}\boldsymbol{Z}_{0,p,N_p}$ can be calculated. This product corresponds, by definition, to the extended observability matrix times the state sequence: $\tilde{\Gamma}^{(f)}\boldsymbol{X}_{p,N_p}$. Using an SVD, the order of the system and the state sequence can then be estimated

$$\tilde{\Gamma}^{(f)}\boldsymbol{X}_{p,N_p} = \tilde{\Gamma}^{(f)}\tilde{\mathcal{K}}^{(p)}\boldsymbol{Z}_{0,p,N_p} = \mathcal{U}_n \Sigma_n \mathcal{V}_n^\top \tag{18}$$

Note that it is possible to include a left weighting matrix in this equation. Such a weight may affect the variance of the estimated state sequence and the resulting identified system. The choice and design of such weightings has been the topic of many discussions, see for example, [2, 24, 59, 60].

The state sequence is recovered (up to a similarity transformation) as

$$\hat{\boldsymbol{X}}_{p,N_p} = \Sigma_n \mathcal{V}_n^\top \tag{19}$$

In practice, the matrix $\tilde{\Gamma}^{(f)}\tilde{\mathcal{K}}^{(p)}$ is constructed using estimated parameters. Hence, the SVD in (18) will not exactly contain $n$ non-zero singular values and we will only obtain an estimate of the true state sequence $\hat{\boldsymbol{X}}_{p,N_p}$. Note that the choice of the future window $f$ receives further attention in Section 3.5.2.

Subsequently, (16) and (17) are solved in a least-squares sense. First, (17) is solved and subsequently its residual $\hat{\boldsymbol{E}}_{p,N_p}$ is used in the solution of (16).

*3.3.1 Variants:* Several variants of the PBSID$_{\text{opt}}$ algorithm can be found in the literature, some of which have already been mentioned. The PBSID$_{\text{opt}}$ is asymptotically equivalent [2] to the SSARX algorithm proposed by Jansson [14]. Several other modifications are discussed in [61].

*3.3.2 Recursive implementation:* The PBSID$_{\text{opt}}$ can be also employed for on-line applications, by working out recursive implementations. The problem has been studied in the literature by a number of authors and a template for recursive closed-loop subspace identification can be outlined as follows, along the lines of the general ideas proposed in [62] and of the algorithm in [63]):

- Recursive update of the solution of the least-squares problem (9), using a conventional recursive least squares (RLS) scheme.
- Update of the estimate of the state sequence, that is, of the state estimate given by (19). In this respect, note that this is the most critical step in the implementation, as one has to ensure that the recursive state estimates are expressed in a consistent state-space basis. One way of guaranteeing this is given by, for example, the scheme proposed in [63], which is based on the so-called propagator method for the recursive update of the state sequence (see also [64] for details).
- Recursive estimate of the state-space matrices of the system, that is, update of the solution of the least-squares problem (16), again by means of RLS.

### 3.4 Closed-loop MOESP

In Section 2.2.3, it was shown that besides the Markov parameters, an estimate of the innovation sequence can be obtained. If we revisit the innovation model (1) an estimate, $\{\hat{\boldsymbol{e}}_k\}_{k=p}^{N-1}$, of the innovation sequence $\{\boldsymbol{e}_k\}_{k=p}^{N-1}$ is now available (although not over the full data horizon $k = 0, \ldots, N-1$). First obtaining an estimate of the innovation sequence results in a class of 'innovation estimation' methods, first introduced by Qin and Ljung and leading to their PARSIM

method [65]. Having knowledge of the innovation sequence, we are effectively left with a deterministic identification problem to which solutions are well-known. The deterministic MOESP algorithm is one such algorithm. The method described here, 67, was presented in [5, 66, 67] and is inspired by and in many respects similar to [3, 4]. In particular, Qin and Ljung [3] suggest a recursive procedure to estimate the innovation sequence and Di Ruscio [4] suggests the same method of estimating the innovation sequence but proceeds using the DSR method (see Section 3.4.4 for some more details).

*3.4.1 Estimating the extended observability matrix $\Gamma^{(f)}$:* Referring to [43] for a detailed derivation, we consider the following data equation in the MOESP algorithm

$$\boldsymbol{Y}_{p,f,N_f} = \Gamma^{(f)}\boldsymbol{X}_{p,N_f} + \boldsymbol{H}^{(f)}(\boldsymbol{B},\boldsymbol{D})\boldsymbol{U}_{p,f,N_f} + \boldsymbol{H}^{(f)}(\boldsymbol{K},0)\boldsymbol{E}_{p,f,N_f}$$

with $N_f = N - f - p + 1$. This is also the data equation considered in the deterministic MOESP setting, since the signals constituting $\boldsymbol{U}_{p,f,N_f}$ and $\boldsymbol{E}_{p,f,N_f}$ are both at our disposal and no further stochastic disturbances are present. The influences of the input and the innovation can be eliminated using orthogonal subspace projection. For this purpose, we construct the orthogonal projection matrix

$$\Pi^{\perp}_{\mathcal{Z}_{p,f,N_f}} = \boldsymbol{I} - \mathcal{Z}^{\dagger}_{p,f,N_f}\mathcal{Z}_{p,f,N_f}$$

where the definition $\mathcal{Z}_{p,f,N_f} = \begin{bmatrix}\boldsymbol{U}_{p,f,N_f}\\ \boldsymbol{E}_{p,f,N_f}\end{bmatrix}$ is used. Applying this projection results in

$$\boldsymbol{Y}_{p,f,N_f}\Pi^{\perp}_{\mathcal{Z}_{p,f,N_f}} = \Gamma^{(f)}\boldsymbol{X}_{p,N_f}\Pi^{\perp}_{\mathcal{Z}_{p,s,N_f}}$$

In practice, this projection may be obtained by performing an RQ factorisation of the input and output data, which is numerically much more efficient and stable than evaluating the large projection matrix

$$\begin{bmatrix}\mathcal{Z}_{p,f,N_f}\\ \boldsymbol{Y}_{p,f,N_f}\end{bmatrix} = \begin{bmatrix}\boldsymbol{R}_{11} & 0\\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{Q}_1\\ \boldsymbol{Q}_2\end{bmatrix}$$

Using the properties of the RQ factorisation, we can then equivalently write

$$\boldsymbol{Y}_{p,f,N_f}\Pi^{\perp}_{\mathcal{Z}_{p,f,N_f}} = \Gamma^{(f)}\boldsymbol{X}_{p,N_f}\Pi^{\perp}_{\mathcal{Z}_{p,f,N_f}} = \boldsymbol{R}_{22}\boldsymbol{Q}_2$$

If the input and noise sequences are persistently exciting of at least order $fn_u$ and $fn_y$, respectively, for the input and innovation signals [43], the following holds, since then the column space of the observability matrix is preserved in $\boldsymbol{Y}_{p,f,N_f}\Pi^{\perp}_{\mathcal{Z}_{p,f,N_f}}$ after projection

$$\begin{aligned}\text{range}(\Gamma^{(f)}) &= \text{range}\left(\lim_{N\to\infty}\frac{1}{\sqrt{N-p}}\boldsymbol{Y}_{p,f,N_f}\Pi^{\perp}_{\mathcal{Z}_{p,f,N_f}}\right)\\ &= \text{range}\left(\lim_{N\to\infty}\frac{1}{\sqrt{N-p}}\boldsymbol{R}_{22}\right)\end{aligned}$$

Thus, due to the fact that $\boldsymbol{Q}_2$ has full row rank, the column-space of $\boldsymbol{R}_{22}$ serves as a basis for the column space of the extended observability matrix $\mathcal{O}_f$. Performing an SVD of $\boldsymbol{R}_{22}$ gives

$$\boldsymbol{R}_{22} = \mathcal{U}_n\Sigma_n\mathcal{V}_n^{\top}$$

where $n$ is the number of dominant singular values and also the order of the underlying innovation system. The columns of $\mathcal{U}_n$ provide a basis for $\Gamma^{(f)}$. A gap between successive singular values will often indicate the order of the system (see [68] for more details). As in the PBSID$_{\text{opt}}$ method, the SVD can be augmented with weights, which may change the variance on the estimated observability matrix, see for example, the 'unifying theorem' in [24, 41, 59, 60].

*3.4.2 Estimating the $\boldsymbol{A}$ and $\boldsymbol{C}$ matrices:* Estimates of the $\boldsymbol{A}$ and $\boldsymbol{C}$ matrices can subsequently be obtained from $\mathcal{U}_n$. Given the structure of $\Gamma^{(f)}$, the $\boldsymbol{C}$ matrix is found as the first $n_y$ rows of $\mathcal{U}_n$. $\boldsymbol{A}$ can be found as the solution to the overdetermined problem

$$\mathcal{U}_n(1\!:\!(f-1)n_y,:)\boldsymbol{A} = \mathcal{U}_n(n_y+1\!:\!fn_y,:) \tag{20}$$

(using MATLAB notation). The matrices $\boldsymbol{B}$, $\boldsymbol{D}$ and $\boldsymbol{K}$ can be computed in a second step by solving a least-squares problem as shown in the next subsection.

*3.4.3 Estimating $\boldsymbol{B}$, $\boldsymbol{D}$, $\boldsymbol{K}$ and the initial state:* Based on the system description (1), the output at time $k$ can be written as

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{A}^k\boldsymbol{x}_0 + \sum_{\tau=0}^{k-1}\boldsymbol{C}\boldsymbol{A}^{k-\tau-1}(\boldsymbol{B}\boldsymbol{u}_\tau + \boldsymbol{K}\boldsymbol{e}_\tau) + \boldsymbol{D}\boldsymbol{u}_k + \boldsymbol{e}_k$$

Applying the vectorisation operator and exploiting a property of the Kronecker product [69], this can be rewritten as

$$\boldsymbol{y}_k = \underbrace{\begin{bmatrix}\boldsymbol{\Phi}_k^{x_0} & \boldsymbol{\Phi}_k^{\boldsymbol{B}} & \boldsymbol{\Phi}_k^{\boldsymbol{K}} & \boldsymbol{\Phi}_k^{\boldsymbol{D}}\end{bmatrix}}_{\boldsymbol{\Phi}_k}\underbrace{\begin{bmatrix}\boldsymbol{x}_0\\ \text{vec}(\boldsymbol{B})\\ \text{vec}(\boldsymbol{K})\\ \text{vec}(\boldsymbol{D})\end{bmatrix}}_{\boldsymbol{\Theta}} + \boldsymbol{e}_k \tag{21}$$

where we have defined

$$\boldsymbol{\Phi}_k^{x_0} = \boldsymbol{C}\boldsymbol{A}^{k-1}, \quad \boldsymbol{\Phi}_k^{\boldsymbol{B}} = \sum_{\tau=0}^{k-1}\boldsymbol{u}_\tau^{\top}\otimes\boldsymbol{C}\boldsymbol{A}^{k-\tau-1}$$

$$\boldsymbol{\Phi}_k^{\boldsymbol{K}} = \sum_{\tau=0}^{k-1}\boldsymbol{e}_\tau^{\top}\otimes\boldsymbol{C}\boldsymbol{A}^{k-\tau-1}, \quad \boldsymbol{\Phi}_k^{\boldsymbol{D}} = \boldsymbol{u}_k^{\top}\otimes\boldsymbol{I}_{n_y}$$

Equation (21) is a linear expression in the unknown elements of $\boldsymbol{x}_0$, $\boldsymbol{B}$, $\boldsymbol{D}$ and $\boldsymbol{K}$, which can be solved for the parameters in a least-squares sense with the available data set $\{\boldsymbol{u}_k,\boldsymbol{y}_k,\boldsymbol{e}_k\}_{k=p}^{N-1}$. For this purpose, define $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Theta}$ as in (21). Then the least-squares problem can be stated as

$$\hat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}}\sum_{k=p}^{N-1}\|\boldsymbol{y}_k - \boldsymbol{\Phi}_k\boldsymbol{\Theta}\|_2^2 = \arg\min_{\boldsymbol{\Theta}}\|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{\Theta}\|_2^2$$

which can be solved efficiently using a QR factorisation [48]. It must be noted that, although the described way to find $\boldsymbol{B}$, $\boldsymbol{D}$, $\boldsymbol{K}$ and the initial state is conceptually simple, it is computationally prohibitive because of the Kronecker products involved. Much more efficient procedures exist to find the matrix $\boldsymbol{\Phi}$, which avoid evaluation of Kronecker products for each $k$. Such procedures are detailed in [67, 70]. The approach presented here must be modified slightly in case the system is open-loop unstable; see [8].




doi: 10.1049/iet-cta.2012.0653

*3.4.4 Variants:* Here, we have considered one possible 'innovation estimation' algorithm, which we have chosen for its simplicity. One variant is the first innovation estimation algorithm: the PARSIM method presented in [3], where the innovation and Markov parameters are estimated in a recursive fashion. Subsequently, the DSR_e algorithm was presented in [71], [72, Chapter 6.2] and [4]. In the DSR_e algorithm the innovation sequence is estimated in the same way as described in this section. The deterministic subspace identification problem that remains is then solved using the DSR algorithm [72–74], which uses a slightly different data equation than the MOESP method. We estimate the extended observability matrix using the deterministic MOESP method [43, 66].

*Remark 3.2:* In the four previous subsections, estimates have been obtained for the matrix $\boldsymbol{K}$ using least-squares techniques. Although these estimates minimise the appropriate least-squares criteria, they do not automatically satisfy the requirement of having a stable predictor and thus are not proper Kalman gains. If the identified models are to be used in the context of observing or prediction, this is a crucial property. An alternative way to obtain $\boldsymbol{K}$ is by solving a discrete algebraic Riccati equation (DARE). Such approaches are described in [40, 41, 43].

### 3.5 User choices and other issues

Having paid attention to the main algorithmic issues in the previous sections, this section focuses on some of the important user choices and related issues in the closed-loop subspace algorithms. Many of these choices and issues have been studied in the literature and it remains an area of active research with several open problems.

*3.5.1 Past window:* Algorithmically, the main effect of the past window size $p$ is to ensure that the neglected term $\tilde{\boldsymbol{A}}^p$ in (5) is so small that it can be neglected. One should keep in mind, however, that the number of parameters estimated in the least-squares problem (9) is $pn_y(n_u + n_y) + n_y n_u$ and hence grows linearly with $p$, which is the order of the VARX model (7). Thus, when using finite-length data sequences the variance will grow and there is a risk of overfitting. Order selection tools such as the Akaike Information Criterion [1] could be used to avoid such issues in selecting the order of the VARX model (7). Cross-validation between data sets is also an option.

It may also be advantageous to employ regularisation in the least-squares regression problem, for example, using Tikhonov regularisation that was pioneered in [75] and has been implemented in [76], or $\ell_1$-regularised regression. The bias because of a finite $p$ can be remedied by using a Vector ARMAX structure, as was proposed in [77]. This approach was shown to work well when the order of the identified system corresponds exactly to the order of the 'true' system.

Asymptotic consistency results, for example, for $p \to \infty$, can be found in [2, 23, 60, 78], while the effect of choosing a finite $p$ remains hard to quantify for general situations. First results on the statistical behaviour of subspace algorithms as a function of the window size $p$ were reported in [60] where the authors study the effect of $p$ on the variance of the estimated system's invariants for white inputs. Chiuso [2] also discusses the effect of $p$ in the context of the recent closed-loop subspace methods (PBSID). It is shown that $p$ must be chosen in relation to the number of available samples $N$ to result in the statistically optimal choice. Finally, Kuersteiner [78] focusses on how the infinite-order ARX model that consistently describes the underlying system (i.e. (5) when letting $p \to \infty$) can be approximated by an ARX model of finite order and how $p$ must be chosen to achieve certain statistical properties.

*3.5.2 Future window:* The choice of the future window in the $\text{PBSID}_{\text{opt}}$ algorithm in Section 3.3, constrained by $f \geq n$, affects the variance of the invariants associated with the identified system (e.g. the elements of the system matrices in a certain fixed basis, pole locations or the transfer function). In particular, it can be shown for certain classes of inputs that the variance on these invariants is a non-increasing function of the future window. These issues are extensively discussed in [46]. Tools to compute the asymptotic variance on the system matrices estimated with the $\text{PBSID}_{\text{opt}}$ method are discussed in [79]. For the choice of the subspace dimension parameters in the CLMOESP algorithm, we refer to results for the classical subspace algorithms [60, 80, 81].

*3.5.3 Incorporating prior knowledge:* In recent years, the possibility of incorporating certain prior knowledge on the system to be identified has regularly received attention. For instance, it is possible to enforce stability or positive realness of the identified models. In [42, 82], it is shown how the least-squares regressions (cf. (15), (16) and (20)) can be modified in a simple way by adding a regularisation term. The amount of regularisation required to achieve a specific spectral radius of $\boldsymbol{A}$, or positive realness of the system can be determined by solving a generalised eigenvalue problem. The former has been implemented in the PBSID toolbox [76]. In [83], it is shown how linear matrix inequalities can be formulated to constrain the eigenvalues of the identified system to lie in certain convex regions. Finally, in [84] some steps are made towards prescribing a certain model structure (e.g. OE, ARMAX, etc.) in subspace methods. Constraints can also be added to incorporate certain prior information on the input–output behaviour, such as the steady-state gain, as discussed in [85, 86]. In a similar vein, research has been directed towards allowing more accurate estimation of finite-order models in cases where this is not trivial, for instance using nuclear norm regularisation [87, 88].

## 4 Continuous-time identification framework

In this section, we consider the continuous-time formulation of the identification problem that was addressed in Section 2 for discrete-time systems. As in the discrete-time case, the underlying assumption is that the continuous-time system to be identified is a finite-dimensional, linear, time-invariant system, subject to measurement and/or process noise. In the continuous-time identification problem, we consider the following continuous-time LTI system

$$\Sigma_{\text{c}}: \begin{cases} \mathrm{d}\boldsymbol{x}(t) = \boldsymbol{A}_{\text{c}}\boldsymbol{x}(t)\,\mathrm{d}t + \boldsymbol{B}_{\text{c}}\boldsymbol{u}(t)\,\mathrm{d}t + \mathrm{d}\boldsymbol{w}(t) & \text{(22a)} \\ \mathrm{d}\boldsymbol{z}(t) = \boldsymbol{C}_{\text{c}}\boldsymbol{x}(t)\,\mathrm{d}t + \boldsymbol{D}_{\text{c}}\boldsymbol{u}(t)\,\mathrm{d}t + \mathrm{d}\boldsymbol{v}(t) & \text{(22b)} \\ \boldsymbol{y}(t)\mathrm{d}t = \mathrm{d}\boldsymbol{z}(t) & \text{(22c)} \end{cases}$$

where $(\boldsymbol{A}_{\text{c}}, \boldsymbol{B}_{\text{c}}, \boldsymbol{C}_{\text{c}}, \boldsymbol{D}_{\text{c}})$ are the continuous-time counterparts of $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$. In the following it will be assumed that $(\boldsymbol{A}_{\text{c}}, \boldsymbol{C}_{\text{c}})$ is observable and that $(\boldsymbol{A}_{\text{c}}, [\boldsymbol{B}_{\text{c}}, \boldsymbol{Q}_{\text{c}}^{1/2}])$ is controllable. The process noise $\boldsymbol{w}(t)$ and measurement noise $\boldsymbol{v}(t)$

are described as Wiener processes with covariance of the increments given by

$$\mathbb{E}\left\{\begin{bmatrix}\mathrm{d}\boldsymbol{w}(t)\\ \mathrm{d}\boldsymbol{v}(t)\end{bmatrix}\begin{bmatrix}\mathrm{d}\boldsymbol{w}(t)\\ \mathrm{d}\boldsymbol{v}(t)\end{bmatrix}^{\top}\right\}=\begin{bmatrix}\boldsymbol{Q} & \boldsymbol{S}\\ \boldsymbol{S}^{\top} & \boldsymbol{R}\end{bmatrix}\mathrm{d}t$$

Note that the definition of the model in terms of increments is because of well-known technical difficulties in the definition of continuous-time white noise processes with finite variance. Formally, such processes do not exist. Hence, the noises are modelled as Wiener processes in terms of independent increments. The reader is referred to [36, Sec. 3.3] and [89, Appendix 1] for more details.

The continuous-time model $\Sigma_{\mathrm{c}}$ also admits an innovation representation given by

$$\mathcal{P}_{\mathrm{c}}:\begin{cases}\mathrm{d}\boldsymbol{x}(t)=\boldsymbol{A}_{\mathrm{c}}\boldsymbol{x}(t)\,\mathrm{d}t+\boldsymbol{B}_{\mathrm{c}}\boldsymbol{u}(t)\,\mathrm{d}t+\boldsymbol{K}_{\mathrm{c}}\mathrm{d}\boldsymbol{e}(t) & \text{(23a)}\\ \mathrm{d}\boldsymbol{z}(t)=\boldsymbol{C}_{\mathrm{c}}\boldsymbol{x}(t)\,\mathrm{d}t+\boldsymbol{D}_{\mathrm{c}}\boldsymbol{u}(t)\,\mathrm{d}t+\mathrm{d}\boldsymbol{e}(t) & \text{(23b)}\\ \boldsymbol{y}(t)\mathrm{d}t=\mathrm{d}\boldsymbol{z}(t) & \text{(23c)}\end{cases}$$

where $\boldsymbol{K}_{\mathrm{c}}\in\mathbb{R}^{n\times n_y}$ is the Kalman gain.

We can now formulate the continuous-time identification problem, which is very similar to the discrete-time version:

*Problem 4.1 Continuous-time subspace identification problem:* Based on a finite set of (not necessarily uniformly) sampled input and output data $\{\boldsymbol{u}_k,\boldsymbol{y}_k\}_{k=0}^{N-1}$ obtained from system $\Sigma_{\mathrm{c}}$, estimate the order $n$ of the continuous-time system $\mathcal{P}_{\mathrm{c}}$ and the associated system matrices $(\boldsymbol{A}_{\mathrm{c}},\boldsymbol{B}_{\mathrm{c}},\boldsymbol{C}_{\mathrm{c}},\boldsymbol{D}_{\mathrm{c}},\boldsymbol{K}_{\mathrm{c}})$ up to a similarity transformation.

### 4.1 Preliminaries

One of the main issues in continuous-time identification is the need to compute high-order derivatives of measured input–output data, which arise from the continuous-time formulations of the data equations used in subspace identification. Obviously, numerical differentiation is not feasible in this context. In [90, 91] the combination of the MOESP algorithm with filtering methods to avoid the need to compute numerical derivatives of input–output signals was proposed. When focusing on the closed-loop case, two approaches have been studied in the literature to circumvent this problem. Both these approaches rely on signal transformations, either through filtering or projections, associated with the class of Laguerre orthogonal filters. We will first introduce the notation and definitions required for the subsequent derivations.

In the following, let $\mathcal{L}_2(0,\infty)$ denote the space of square integrable and Lebesgue measurable functions with inner product $\langle f,g\rangle=\int_0^\infty f(t)g(t)\,\mathrm{d}t$, for $f,g\in\mathcal{L}_2(0,\infty)$. The space $\mathcal{H}_2$ is the closed subspace of $\mathcal{L}_2(j\mathbb{R})$ with functions analytic in the open right half plane. The spaces $\mathcal{L}_2(0,\infty)$ and $\mathcal{H}_2$ are related by the Fourier transform, that is, if $U\in\mathcal{H}_2$, $u=\mathcal{F}^{-1}\{U\}$ is in $\mathcal{L}_2(0,\infty)$ and vice versa. A scalar function $w(s)$ is called inner if $w(j\omega)\in\mathcal{H}_2$, such that $|w(j\omega)|=1$ almost everywhere on the imaginary axis. Finally, we denote by $\Lambda_w$ the multiplication operator $\mathcal{L}_2(0,\infty)\mapsto\mathcal{L}_2(0,\infty)$ defined by

$$\Lambda_w\boldsymbol{u}(t)=\mathcal{F}^{-1}\{w\mathcal{F}\{\boldsymbol{u}(t)\}\}$$

In the following, we make use of the first-order inner transfer function

$$w(s)=\frac{s-a}{s+a}\qquad(24)$$

with $a>0$. It can be shown that $w(s)\mathcal{H}_2$ is a proper closed subspace of $\mathcal{H}_2$, the orthogonal complement of which is denoted as $\mathcal{S}=\mathcal{H}_2\ominus w(s)\mathcal{H}_2$, that

$$\mathcal{L}_0(s)=\frac{c_w}{s+a}=\frac{\sqrt{2a}}{s+a}$$

is a basis of the one-dimensional subspace $\mathcal{S}$ and that the set

$$\{\mathcal{L}_0,\ \mathcal{L}_0w,\ldots,\mathcal{L}_0w^i,\ldots\}$$

is an orthonormal basis for $\mathcal{H}_2$. Equivalently, letting $\ell_0=\mathcal{F}^{-1}\{\mathcal{L}_0\}$ (i.e. $\ell_0$ is the impulse response of $\mathcal{L}_0$), the set

$$\{\ell_0,\Lambda_w\ell_0,\ldots,\Lambda_w^i\ell_0,\ldots\}\qquad(25)$$

is an orthonormal basis for $\mathcal{L}_2(0,\infty)$.

Making use of these definitions, the transfer function of the $i$th Laguerre filter (of order $i+1$) may be defined as

$$\mathcal{L}_i(s)=w^i(s)\mathcal{L}_0(s)=\sqrt{2a}\frac{(s-a)^i}{(s+a)^{i+1}}\qquad(26)$$

In accordance with [36], we will also define the Laguerre-like filters

$$L_i(s)=2a\frac{(s-a)^i}{(s+a)^{i+1}}=(1-w)w^i\qquad(27)$$

These Laguerre-like filters are introduced for notational simplicity and can be used since the normalising constant $\sqrt{2a}$ is immaterial to the subsequent results.

In the time domain, we denote with $w^i(t)$ the impulse response of the concatenation of $i$ all-pass filters (24) and by $[w^i\boldsymbol{u}](t)$ the convolution of $\boldsymbol{u}(t)$ and $w^i(t)$, that is, $[w^i\boldsymbol{u}](t)\equiv\int_0^t w^i(t-\tau)\boldsymbol{u}(\tau)\mathrm{d}\tau$. Similarly, $l_i(t)$ is the impulse response of the $i$th Laguerre-like filter (27) and $[l_i\boldsymbol{u}](t)$ denotes the convolution of $\boldsymbol{u}(t)$ and $l_i(t)$, that is, $[l_i\boldsymbol{u}](t)\equiv\int_0^t l_i(t-\tau)\boldsymbol{u}(\tau)\,\mathrm{d}\tau$.

### 4.2 A Laguerre filtering approach

The first approach relies on the idea, presented in [36, 92], of resorting to the bilinear transformation associated with a first-order all-pass filter in order to convert the derivative operation into a low-pass filtering operation based on the class of Laguerre filters. It can be shown that a data equation can be derived, which is similar to the one for discrete-time systems and hence subspace identification techniques can be applied.

On the basis of the definitions in the previous subsection, the innovation model (23) can be transformed to the 'all-pass' domain by means of the Laguerre-like filters [36, Lemma 3.4], resulting in the equivalent model

$$\begin{aligned}[w\boldsymbol{x}](t)&=\boldsymbol{A}_w\boldsymbol{x}(t)+\boldsymbol{B}_w[l_0\boldsymbol{u}](t)+\boldsymbol{K}_w[l_0\boldsymbol{e}](t)+\boldsymbol{F}_1\boldsymbol{x}_0l_0(t)\\ [l_0\boldsymbol{y}](t)&=\boldsymbol{C}_w\boldsymbol{x}(t)+\boldsymbol{D}_w[l_0\boldsymbol{u}](t)+[l_0\boldsymbol{e}](t)+\boldsymbol{F}_2\boldsymbol{x}_0l_0(t)\end{aligned}$$

where the state-space matrices are given by

$$\begin{aligned}\boldsymbol{A}_w&=(\boldsymbol{A}_{\mathrm{c}}+a\boldsymbol{I})^{-1}(\boldsymbol{A}_{\mathrm{c}}-a\boldsymbol{I})\\ \boldsymbol{B}_w&=(\boldsymbol{A}_{\mathrm{c}}+a\boldsymbol{I})^{-1}\boldsymbol{B}_{\mathrm{c}}\\ \boldsymbol{C}_w&=2a\boldsymbol{C}_{\mathrm{c}}(\boldsymbol{A}_{\mathrm{c}}+a\boldsymbol{I})^{-1}\\ \boldsymbol{D}_w&=\boldsymbol{D}_{\mathrm{c}}-\boldsymbol{C}_{\mathrm{c}}(\boldsymbol{A}_{\mathrm{c}}+a\boldsymbol{I})^{-1}\boldsymbol{B}_{\mathrm{c}}\\ \boldsymbol{F}_1&=(\boldsymbol{A}_{\mathrm{c}}+a\boldsymbol{I})^{-1}\\ \boldsymbol{F}_2&=\boldsymbol{C}_{\mathrm{c}}(\boldsymbol{A}_{\mathrm{c}}+a\boldsymbol{I})^{-1}\end{aligned}\qquad(28)$$




doi: 10.1049/iet-cta.2012.0653

and $\boldsymbol{x}_0$ is the initial state of the original continuous-time system. Just as for the original system, $[l_0\boldsymbol{e}](t)$ may be eliminated from the state equation to result in a one-step-ahead predictor

$$[w\boldsymbol{x}](t) = \tilde{\boldsymbol{A}}_w\boldsymbol{x}(t) + \tilde{\boldsymbol{B}}_w[l_0\boldsymbol{u}](t) + \boldsymbol{K}_w[l_0\boldsymbol{y}](t) + \tilde{\boldsymbol{F}}_w\boldsymbol{x}_0 l_0(t), \tag{29a}$$

$$[l_0\boldsymbol{y}](t) = \boldsymbol{C}_w\boldsymbol{x}(t) + \boldsymbol{D}_w[l_0\boldsymbol{u}](t) + [l_0\boldsymbol{e}](t) + \boldsymbol{F}_2\boldsymbol{x}_0 l_0(t), \tag{29b}$$

where we have introduced $\tilde{\boldsymbol{A}}_w \equiv \boldsymbol{A}_w - \boldsymbol{K}_w\boldsymbol{C}_w$, $\tilde{\boldsymbol{B}}_w \equiv \boldsymbol{B}_w - \boldsymbol{K}_w\boldsymbol{D}_w$ and $\tilde{\boldsymbol{F}}_w \equiv \boldsymbol{F}_1 - \boldsymbol{K}_w\boldsymbol{F}_2$.

Considering system (29) in the all-pass domain, we can follow an approach similar to the one in Section 2.2 for the discrete-time problem to estimate the Markov parameters of (29). We first redefine some of the matrices defined before in Section 2.1. We introduce a stacked sample of input and output data $[l_i\boldsymbol{z}](t)$ according to

$$[l_i\boldsymbol{z}](t) = \begin{bmatrix} [l_i\boldsymbol{u}](t) \\ [l_i\boldsymbol{y}](t) \\ l_{i,t} \end{bmatrix}$$

The stacked vector $\boldsymbol{z}_i^{(p)}(t)$ is defined as

$$\boldsymbol{z}_i^{(p)}(t) = \left[[l_{i-p}\boldsymbol{z}](t)^\top, \quad [l_{i-p+1}\boldsymbol{z}](t)^\top, \ldots, [l_{i-1}\boldsymbol{z}](t)^\top\right]^\top$$

where $p$ denotes the 'past window' size. Note that the indices $\{i-p, \ldots, i-1\}$, which corresponded to time shifts in the discrete-time case, now denote repeated filtering operations on the input–output data.

We also define a reversed extended controllability matrix $\tilde{\mathcal{K}}^{(p)}$

$$\tilde{\mathcal{K}}_w^{(p)} = \left[\tilde{\boldsymbol{A}}_w^{p-1}\bar{\boldsymbol{B}}_w, \quad \tilde{\boldsymbol{A}}_w^{p-2}\bar{\boldsymbol{B}}_w, \ldots, \bar{\boldsymbol{B}}_w\right] \tag{30}$$

where we have defined $\bar{\boldsymbol{B}}_w = [\tilde{\boldsymbol{B}}_w, \ \boldsymbol{K}_w, \ \tilde{\boldsymbol{F}}_w\boldsymbol{x}_0]$ for brevity. Starting from some initial state $[w\boldsymbol{x}](t)$, the state equation of (30) can be iterated, resulting in the expression

$$[w^p\boldsymbol{x}](t) = \tilde{\boldsymbol{A}}_w^p\boldsymbol{x}(t) + \tilde{\mathcal{K}}_w^{(p)}\boldsymbol{z}_p^{(p)}(t)$$

Thus, under the assumption that $\tilde{\boldsymbol{A}}_w$ is stable, we can express the outputs of the system in terms of repeatedly filtered input–output data as

$$[l_p\boldsymbol{y}](t) = \boldsymbol{C}_w\tilde{\mathcal{K}}_w^{(p)}\boldsymbol{z}_p^{(p)}(t) + \boldsymbol{D}_w[l_p\boldsymbol{u}](t) + \boldsymbol{F}_2\boldsymbol{x}_0 l_p(t) + [l_p\boldsymbol{e}](t) \tag{31}$$

*Remark 4.1:* Note that, as in the discrete-time algorithm, a shift invariance property can be exploited to obtain shifted versions of this equation. That is, we may obtain

$$\begin{aligned}[l_{p+j-1}\boldsymbol{y}](t) = {} & \boldsymbol{C}_w\tilde{\mathcal{K}}_w^{(p)}\boldsymbol{z}_{p+j-1}^{(p)}(t) + \boldsymbol{D}_w[l_{p+j-1}\boldsymbol{u}](t) \\ & + \boldsymbol{F}_2\boldsymbol{x}_0 l_{p+j-1}(t) + [l_{p+j-1}\boldsymbol{e}](t)\end{aligned} \tag{32}$$

for an arbitrary $j$. In this discussion, we will not use these shifted versions and the reader is referred to [35] for details on how these can be exploited.

Considering now the (not necessarily uniformly) sampled input–output data of the system at instants $t \in \{t_0, t_1, \ldots, t_{N-1}\}$ and realising that (31) holds for all the sample instants, it follows we can estimate the parameters $\boldsymbol{C}_w\tilde{\mathcal{K}}_w^{(p)}$ and $\boldsymbol{D}_w$ in a least-squares problem. For this purpose, we define

$$\boldsymbol{Z}_{0,p,N}^w = \left[\boldsymbol{z}_p^{(p)}(t_0), \quad \boldsymbol{z}_p^{(p)}(t_1), \ldots, \boldsymbol{z}_p^{(p)}(t_{N-1})\right]$$

and

$$\boldsymbol{Y}_{p,N}^w = \left[[l_p\boldsymbol{y}](t_0), \quad [l_p\boldsymbol{y}](t_1), \ldots, [l_p\boldsymbol{y}](t_{N-1})\right]$$

where the subscript $p$ indicates filtering with the $p$th Laguerre-like filter and the subscript $N$ indicates the number of columns. The matrices $\boldsymbol{U}_{p,N}^w$, $\boldsymbol{E}_{p,N}^w$, $\boldsymbol{\Psi}_{p,N}$ are defined likewise. Then, we can formulate the data equation

$$\boldsymbol{Y}_{p,N}^w = \boldsymbol{C}_w\tilde{\mathcal{K}}_w^{(p)}\boldsymbol{Z}_{0,p,N}^w + \boldsymbol{D}_w\boldsymbol{U}_{p,N}^w + \boldsymbol{F}_2\boldsymbol{x}_0\boldsymbol{\Psi}_{p,N} + \boldsymbol{E}_{p,N}^w$$

Based on this data equation, we may obtain least-squares estimates of $\boldsymbol{C}_w\tilde{\mathcal{K}}_w^{(p)}$ and $\boldsymbol{D}_w$.

*4.2.1 Continuous-time PBSID$_{\text{opt}}$:* Using the estimated Markov parameters $\widehat{\boldsymbol{C}_w\tilde{\mathcal{K}}_w^{(p)}}$, we can construct a predictor for the state in a very similar way as for the discrete-time PBSID$_{\text{opt}}$ algorithm; see Section 3.3. To this end, we form the extended observability-times-controllability matrix

$$\tilde{\Gamma}_w^{(f)}\tilde{\mathcal{K}}_w^{(p)} = \begin{bmatrix} \boldsymbol{C}_w\tilde{\boldsymbol{A}}_w^{p-1}\bar{\boldsymbol{B}}_w & \boldsymbol{C}_w\tilde{\boldsymbol{A}}_w^{p-2}\bar{\boldsymbol{B}}_w & \cdots & \boldsymbol{C}_w\bar{\boldsymbol{B}}_w \\ 0 & \boldsymbol{C}_w\tilde{\boldsymbol{A}}_w^{p-1}\bar{\boldsymbol{B}}_w & \cdots & \boldsymbol{C}_w\tilde{\boldsymbol{A}}_w\bar{\boldsymbol{B}}_w \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & & \boldsymbol{C}_w\tilde{\boldsymbol{A}}_w^{f-1}\bar{\boldsymbol{B}}_w \end{bmatrix}$$

From the previous derivations, it follows that the product

$$\tilde{\Gamma}_w^{(f)}\tilde{\mathcal{K}}_w^{(p)}\boldsymbol{Z}_{0,p,N}^w \equiv \tilde{\Gamma}_w^{(f)}\boldsymbol{X}_{p,N}^w \simeq \mathcal{U}_n\Sigma_n\mathcal{V}_n^\top$$

represents, by definition, the product of the extended observability matrix of the all-pass domain model (29) and the state sequence and hence we can estimate the state sequence by means of an SVD

$$\hat{\boldsymbol{X}}_{p,N}^w \simeq \Sigma_n\mathcal{V}_n^\top$$

Note, in this context, that $\boldsymbol{X}_{p,N}^w$ is defined as

$$\boldsymbol{X}_{p,N}^w = \left[[w_p\boldsymbol{x}](t_0), [w_p\boldsymbol{x}](t_1), \ldots, [w_p\boldsymbol{x}](t_{N-1})\right]$$

In a similar way, we can estimate the state sequence $\boldsymbol{X}_{p+1,N}^w$ from the SVD

$$\tilde{\Gamma}_w^{(f)}\tilde{\mathcal{K}}_w^{(p)}\boldsymbol{Z}_{1,p+1,N}^w \equiv \tilde{\Gamma}_w^{(f)}\boldsymbol{X}_{p+1,N}^w \simeq \mathcal{U}_n\Sigma_n\mathcal{V}_n^\top$$

Having estimated the state sequences $\boldsymbol{X}_{p,N}^w$ and $\boldsymbol{X}_{p+1,N}^w$, the state-space matrices $(\boldsymbol{A}_w, \boldsymbol{B}_w, \boldsymbol{C}_w, \boldsymbol{D}_w, \boldsymbol{K}_w)$ can be estimated in a least-squares sense, making use of the relations

$$\boldsymbol{X}_{p+1,N}^w = \begin{bmatrix}\boldsymbol{A}_w & \boldsymbol{B}_w & \boldsymbol{K}_w\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{p,N}^w \\ \boldsymbol{U}_{p,N}^w \\ \boldsymbol{E}_{p,N}^w\end{bmatrix}$$

$$\boldsymbol{Y}_{p,N}^w = \begin{bmatrix}\boldsymbol{C}_w & \boldsymbol{D}_w\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{p,N}^w \\ \boldsymbol{U}_{p,N}^w\end{bmatrix} + \boldsymbol{E}_{p,N}^w$$

Those matrices can be converted to the continuous-time state-space matrices $(\boldsymbol{A}_c, \boldsymbol{B}_c, \boldsymbol{C}_c, \boldsymbol{D}_c, \boldsymbol{K}_c)$ by means of the relations in (28).

*4.2.2 Continuous-time closed-loop MOESP:* Just as for the discrete-time algorithm, we can exploit the knowledge of the estimated innovation matrix $\boldsymbol{E}_{0,N}$ and apply the ordinary MOESP method, but now for continuous-time systems. In the derivation presented here, we use the Laguerre filtering approach. It was shown in [36] that we can formulate an all-pass domain data equation on the basis of the all-pass domain model (29)

$$\begin{aligned}\boldsymbol{Y}^w_{0,f,N} = \Gamma^{(f)}_w \boldsymbol{X}^w_{0,N} &+ \boldsymbol{H}^{(f)}(\boldsymbol{B}_w, \boldsymbol{D}_w)\boldsymbol{U}^w_{0,f,N} \\ &+ \boldsymbol{H}^{(f)}(\boldsymbol{K}_w, \boldsymbol{0})\boldsymbol{E}^w_{0,f,N} \\ &+ \boldsymbol{H}^{(f)}(\boldsymbol{K}_1\boldsymbol{x}_0, \boldsymbol{K}_2\boldsymbol{x}_0)\boldsymbol{\Psi}_{0,f,N}\end{aligned}$$

Just as in the case of the discrete-time CLMOESP algorithm, we can use orthogonal subspace projection to eliminate the influence of the inputs, noise and filtered initial states.

To this end, we compute the RQ factorisation

$$\begin{bmatrix}\begin{bmatrix}\boldsymbol{U}^w_{0,N}\\ \boldsymbol{E}^w_{0,N}\\ \boldsymbol{\Psi}_{0,N}\end{bmatrix}\\ \boldsymbol{Y}^w_{0,N}\end{bmatrix} = \begin{bmatrix}\boldsymbol{R}_{11} & 0\\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{Q}_1\\ \boldsymbol{Q}_2\end{bmatrix}$$

Using the properties of the RQ factorisation, we can then equivalently write

$$\boldsymbol{Y}^w_{0,N}\boldsymbol{Q}_2^\top = \Gamma^{(f)}_w \boldsymbol{X}^w_{0,N}\boldsymbol{Q}_2^\top$$

If the filtered input and noise sequences are persistently exciting of at least order $fn_u$ and $fn_y$, respectively, for the input and innovation signals [43], the following holds

$$\text{range}(\Gamma^{(f)}_w) = \text{range}\left(\lim_{N\to\infty}\frac{1}{\sqrt{N}}\boldsymbol{R}_{22}\right)$$

and hence we can estimate the extended observability matrix of the all-pass domain model from which, in turn, $\boldsymbol{A}_w$ and $\boldsymbol{C}_w$ can be estimated. A least-squares problem can be formulated to estimate the remaining system matrices $\boldsymbol{B}_w$, $\boldsymbol{D}_w$ and $\boldsymbol{K}_w$. We refer the reader to [36, Section 3.7] for details.

### 4.3 A Laguerre projection approach

A second approach is based on the results in [93, 94]. Based on the results in [93], the continuous-time innovation model (23) can be converted to a discrete-time equivalent model as follows. Apply to the input, output and innovation signals of (23) the transformations

$$\begin{aligned}\tilde{\boldsymbol{u}}(k) &= \int_0^\infty \Lambda^k_w \ell_0(t)\boldsymbol{u}(t)\,\mathrm{d}t\\ \tilde{\boldsymbol{y}}(k) &= \int_0^\infty \Lambda^k_w \ell_0(t)\boldsymbol{y}(t)\,\mathrm{d}t\\ \tilde{\boldsymbol{e}}(k) &= \int_0^\infty \Lambda^k_w \ell_0(t)\,\mathrm{d}\boldsymbol{e}(t)\end{aligned}$$

These transformations project the signals onto the $k$th element $\Lambda^k_w\ell_0(t)$ of the orthonormal Laguerre basis (25). Then, it can be shown [93] that the transformed system has the state-space representation

$$\begin{aligned}\boldsymbol{\xi}(k+1) &= \boldsymbol{A}_o\boldsymbol{\xi}(k) + \boldsymbol{B}_o\tilde{\boldsymbol{u}}(k) + \boldsymbol{K}_o\tilde{\boldsymbol{e}}(k)\\ \tilde{\boldsymbol{y}}(k) &= \boldsymbol{C}_o\boldsymbol{\xi}(k) + \boldsymbol{D}_o\tilde{\boldsymbol{u}}(k) + \tilde{\boldsymbol{e}}(k)\end{aligned}$$

where the state-space matrices are given by

$$\begin{aligned}\boldsymbol{A}_o &= (\boldsymbol{A}_\mathrm{c} - a\boldsymbol{I})^{-1}(\boldsymbol{A}_c + a\boldsymbol{I})\\ \boldsymbol{B}_o &= \sqrt{2a}(\boldsymbol{A}_\mathrm{c} - a\boldsymbol{I})^{-1}\boldsymbol{B}_\mathrm{c}\\ \boldsymbol{C}_o &= -\sqrt{2a}\boldsymbol{C}_\mathrm{c}(\boldsymbol{A}_\mathrm{c} - a\boldsymbol{I})^{-1}\\ \boldsymbol{D}_o &= \boldsymbol{D}_\mathrm{c} - \boldsymbol{C}_\mathrm{c}(\boldsymbol{A}_\mathrm{c} - a\boldsymbol{I})^{-1}\boldsymbol{B}_\mathrm{c}\end{aligned} \tag{33}$$

Since the following steps are to a large extent consistent with the steps followed in the Laguerre filtering approach we only give a sketch of the procedure for the Laguerre projection approach. Considering the sampled input–output data of the system at instants $t \in \{t_0,\ t_1, \ldots, t_{N-1}\}$, we apply the following transformations

$$\begin{aligned}\tilde{\boldsymbol{u}}_i(k) &= \int_0^\infty \left(\Lambda^k_w\ell_0(\tau)\right)\boldsymbol{u}(t_i+\tau)\,\mathrm{d}\tau\\ \tilde{\boldsymbol{y}}_i(k) &= \int_0^\infty \left(\Lambda^k_w\ell_0(\tau)\right)\boldsymbol{y}(t_i+\tau)\,\mathrm{d}\tau\\ \tilde{\boldsymbol{e}}_i(k) &= \int_0^\infty \left(\Lambda^k_w\ell_0(\tau)\right)\mathrm{d}\boldsymbol{e}(t_i+\tau)\end{aligned} \tag{34}$$

The transformed system then has the state-space representation

$$\begin{aligned}\boldsymbol{\xi}_i(k+1) &= \boldsymbol{A}_o\boldsymbol{\xi}_i(k) + \boldsymbol{B}_o\tilde{\boldsymbol{u}}_i(k) + \boldsymbol{K}_o\tilde{\boldsymbol{e}}_i(k), \quad \boldsymbol{\xi}_i(0) = \boldsymbol{x}(t_i)\\ \tilde{\boldsymbol{y}}_i(k) &= \boldsymbol{C}_o\boldsymbol{\xi}_i(k) + \boldsymbol{D}_o\tilde{\boldsymbol{u}}_i(k) + \tilde{\boldsymbol{e}}_i(k)\end{aligned} \tag{35}$$

If we now define

$$\begin{aligned}\tilde{\boldsymbol{z}}_i(k) &= \begin{bmatrix}\tilde{\boldsymbol{u}}_i(k)\\ \tilde{\boldsymbol{y}}_i(k)\end{bmatrix}\\ \tilde{\boldsymbol{A}}_o &= \boldsymbol{A}_o - \boldsymbol{K}_o\boldsymbol{C}_o\\ \tilde{\boldsymbol{B}}_o &= \boldsymbol{B}_o - \boldsymbol{K}_o\boldsymbol{D}_o\end{aligned}$$

and $\bar{\boldsymbol{B}}_o = [\tilde{\boldsymbol{B}}_o,\ \boldsymbol{K}_o]$, we can develop a very similar algorithm as in Section 2.2 to compute estimates of the Markov parameters of (35). Subsequently, we can estimate the system matrices $(\boldsymbol{A}_o, \boldsymbol{B}_o, \boldsymbol{C}_o, \boldsymbol{D}_o, \boldsymbol{K}_o)$ using the $\text{PBSID}_{\text{opt}}$ approach, which can subsequently be converted to the continuous-time system matrices by means of the relations (33). For details the reader is referred to [35, 93].

*Remark 4.2:* As remarked in [95] the above described projections based on Laguerre basis functions can only be computed in an approximate sense as the indefinite integrals in (34) need to be truncated to finite intervals. In order to circumvent this difficulty, a novel set of basis functions, for which it can be proved that they are compactly supported, has been proposed in the same paper, where it has been shown that for suitably chosen scalars $\alpha_0$, $\alpha_1$, …, $\alpha_\rho$ (see




doi: 10.1049/iet-cta.2012.0653

[95] for details) the functions

$$\tilde{\ell}_k(t) = \sum_{i=0}^{\rho} \alpha_i \Lambda_w^k \ell_0(t - i\tau) \tag{36}$$

can be used to define the system transformations

$$\begin{aligned} \tilde{\boldsymbol{u}}(k) &= \int_0^\infty \tilde{\ell}_k(t)\boldsymbol{u}(t)\,\mathrm{d}t \\ \tilde{\boldsymbol{y}}(k) &= \int_0^\infty \tilde{\ell}_k(t)\boldsymbol{y}(t)\,\mathrm{d}t \\ \tilde{\boldsymbol{e}}(k) &= \int_0^\infty \tilde{\ell}_k(t)\,\mathrm{d}\boldsymbol{e}(t) \end{aligned} \tag{37}$$

with the property that the signals obtained with the transformations (37) satisfy the discrete-time system (35) for suitably chosen initial states. The support of the functions is compact if $\rho > k$.

### 4.4 Implementation issues

Some additional comments related to the implementation and the choice of the main parameters are in order.

*4.4.1 Impact of discretisation:* First note that the identification algorithms outlined in the previous sections assume that continuous-time filtering of the input–output variables can actually be performed (Section 4.2) or that continuous-time projections of the input–output data on the Laguerre basis can be computed exactly (section 4.3). This is obviously not the case, so suitable discretisation schemes for the considered Laguerre filters and projection operations must be devised.

An analysis of the impact of filter discretisation on the performance of continuous-time SMI algorithms has been carried out in [36]. In particular, it has been shown that a bias term in the estimation of the state-space matrices appears, that is, the discrete-time implementation of the filters leads to a perturbation of the estimated column space of the observability matrix of the system, which in turn leads to bias in the estimated state-space matrices. The perturbation depends on the choice of sampling interval (faster sampling implies smaller bias) and on the conditioning of the system under study (the computed upper bound in the perturbation is inversely proportional to the smallest 'system' singular value in the decomposition leading to the estimate of the observability subspace). Similar issues are expected to arise in the operation of the algorithm proposed in Section 4.2; however, a detailed analysis of the effect of filter discretisation on the bias of the computed estimates of the state-space matrices is not yet available in the literature. As reported in [35], the Laguerre filters used in the implementation of the algorithm can be implemented in discrete-time by means of a conventional bilinear (Tustin) transformation. Similar comments apply to the data projections on the Laguerre basis adopted in the derivation of the algorithm of Section 4.3. In [35], the approximate implementation

$$\begin{aligned} \tilde{\boldsymbol{u}}_i(k) &= \int_0^\infty \Lambda_w^k \ell_0(\tau)\boldsymbol{u}(t_i + \tau)\,\mathrm{d}\tau = \int_{t_i}^\infty \Lambda_w^k \ell_0(\tau - t_i)\boldsymbol{u}(\tau)\,\mathrm{d}\tau \\ &\simeq \int_{t_i}^{t_{N/2}+t_i} \Lambda_w^k \ell_0(\tau - t_i)\boldsymbol{u}(\tau)\,\mathrm{d}\tau \end{aligned} \tag{38}$$

and similarly for $\tilde{\boldsymbol{y}}_i(k)$, has been proposed. In this implementation the indefinite integral has to be computed over a sliding window of length equal to half of the duration of the available dataset.

*4.4.2 Choice of parameters:* Finally, concerning the choice of the parameters $f$ and $p$, the same guidelines as in the case of the discrete-time PBSID algorithm apply. First of all, it is common practice to choose $f = p$, for the sake of simplicity. Then, one has to keep in mind that in view of the need to construct the extended controllability and observability matrices, one has to ensure that $p \geq n$, as a minimum. In addition, $p$ must be sufficiently large to ensure that the term $\bar{\boldsymbol{A}}_w^p \boldsymbol{x}(t)$ is negligible (unfortunately, this can be verified 'a posteriori' only). On the other hand, specific constraints on the choice of $p$ arise because of the need of performing the filtering/projection operations over finite datasets using filters of order up to $2p$. Indeed, it is easy to verify that the impulse responses of the Laguerre and Laguerre-like filters defined in (26) and (27) have a settling time, which is an increasing function of the filter order $k$. Since the implementation of both algorithms requires the computation of convolutions/correlations with the impulse response of such filters over time horizons determined by the length of the available dataset, the duration of the experiment provides an upper bound for the maximum value of the settling time of the highest order filter. An approximate value for the settling time of the filter

$$\mathcal{L}_k(s) = \sqrt{2a}\frac{(s-a)^k}{(s+a)^{k+1}}$$

can be written as $(5 + 2k)/a$, which for $k = 2p$ gives $(5 + 4p)/a$. Therefore denoting with $T$ the duration of the available dataset and letting $\tau = 1/a$, one has the rough guideline $p \leq (aT - 5)/4 \simeq (1/4)(T/\tau) - 1$ for the filtering approach and $p \leq (aT/2 - 5)/4 \simeq (1/8)(T/\tau - 1)$ for the projection approach (the latter in view of the above-described implementation (38) for the projection operation).

*4.4.3 Recursive implementation:* In discrete-time RSMI schemes the recursion is implemented directly with respect to the new discrete input–output sample acquired at the current sampling time. When dealing with the corresponding continuous-time counterpart the first step to be carried out is the (approximate) computation of the projections described in Section 4.3. In this respect, since the index $k$ in the transformed system represents the order of the basis function on which the data has been projected, while the index $i$ is related to the sampling instants $t_i$, the arrival of a new input–output sample leads to the addition of a new time instant at which the projections have to be computed, which leads, in turn, to a new column to be added to the data matrices. In order to compute the projections when dealing with the conventional (i.e. with infinite support) Laguerre filters, the following approximation is introduced

$$\begin{aligned} \tilde{\boldsymbol{u}}_i(k) &= \int_0^\infty \Lambda_w^k \ell_0(\tau)\boldsymbol{u}(t_i + \tau)\,\mathrm{d}\tau \\ &= \int_{t_i}^\infty \Lambda_w^k \ell_0(\tau - t_i)\boldsymbol{u}(\tau)\,\mathrm{d}\tau \\ &\simeq \int_{t_i}^{t_F+t_i} \Lambda_w^k \ell_0(\tau - t_i)\boldsymbol{u}(\tau)\,\mathrm{d}\tau \end{aligned}$$

where $t_F$ is the instant where the impulse response of the filter of maximum order can be considered approximately

equal to zero. In other words

$$t_F = \arg\max_t t$$
$$\text{s.t. } \Lambda_w^{2p-1}\ell_0(t) \le \epsilon$$

where $\epsilon$ is a sufficiently small number. On the other hand, the approximation can be avoided by modifying the basis functions as described in Remark 4.2. In particular, the modified functions have compact support if $\rho > k$; therefore in the following it will be assumed that $\rho > p+f$.

As far as the actual implementation of the recursive $\text{PBSID}_{\text{opt}}$ algorithm is concerned, the approach outlined in Section 3.3.2 can be followed; see [96] for a discussion of the overall recursive algorithm.

## 5 Evaluation

In this section, the closed-loop subspace identification methods discussed in the previous sections are applied to a number of examples. First, we consider two numerical examples. The first example deals with a number of simple closed-loop systems with different characteristics and the second example discusses the differences between $\text{PBSID}_{\text{opt}}$ and CLMOESP in the case of ill-conditioned data because of poor excitation. Next, we consider two experimental examples. In the third example, we evaluate the performance of the main discrete-time algorithms on datasets obtained from a flexible beam and in the fourth example we identify continuous-time models of this same system.

### 5.1 Numerical example: simple closed-loop configurations

In this example, we apply the different methods to a series of closed-loop identification problems taken from [23]. The closed-loop system shown in Fig. 3 is simulated with the transfer functions from Table 1. $e_1$ and $e_2$ are unit variance zero-mean white signals. Second-order models are identified using 1000 samples. For each of the systems the Akaike information criterion in the VARX step was evaluated for a range of $p$. Based on these results, $p = 4$ was suggested for each system and $f = p = 4$ was used subsequently. Each system is identified for 1000 Monte Carlo simulations using the four different methods. The distributions of the identified pole locations, transfer functions and variance-accounted-for (VAF) ($\text{VAF} = \max\{0, (1 - (\text{var}(y - \hat{y})/\text{var}(y))) \times 100\%\}$) [43] on validation data were studied for the combinations in Table 1. It is important to note that these systems are first order, single-input–single-output systems. Therefore they may not highlight some of the aspects that distinguish the different methods when they are applied to MIMO systems with an order that is not clearly defined.

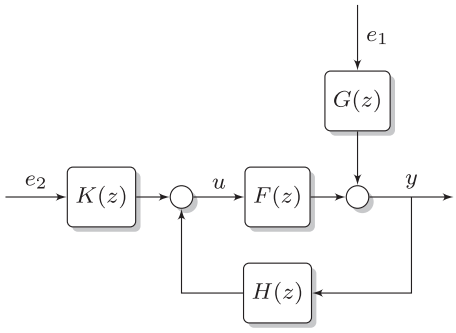


**Fig. 3** *Closed-loop identification setting from [23]*

**Table 1** Systems defined in [23]

| Ex. | $F(z)$ | $H(z)$ | $G(z)$ | $K(z)$ |
|---|---|---|---|---|
| 1 | $\frac{0.3}{z-0.7}$ | $-1$ | $\frac{z+0.5}{z}$ | 1 |
| 2 | $\frac{2.5}{z-3}$ | $-1$ | $\frac{z+0.5}{z}$ | 1 |
| 3 | $\frac{2.5}{z-3}$ | $-1$ | $\frac{z+0.999}{z}$ | $\frac{0.2(z+0.999)}{z-0.99}$ |
| 4 | $\frac{2.5}{z-3}$ | $-1$ | $\frac{z+0.999}{z}$ | 1 |

**Table 2** Mean and standard deviation of the VAF on validation data for the four examples and four methods over 1000 experiments

| Ex. | | Direct | CLMOESP | $\text{PBSID}_{\text{opt}}$ | OKID |
|---|---|---|---|---|---|
| 1 | mean | 93% | 97% | 97% | 97% |
| | std | 3.7% | 2.6% | 2.5% | 2.7% |
| 2 | mean | 99% | 100% | 100% | 100% |
| | std | 0.5% | 0.4% | 0.4% | 0.4% |
| 3 | mean | 100% | 67% | 100% | 96% |
| | std | 0.1% | 43% | 0.1% | 6.8% |
| 4 | mean | 98% | 99% | 99% | 100% |
| | std | 1.3% | 0.6% | 1.7% | 0.5% |

Of the closed-loop methods, $\text{PBSID}_{\text{opt}}$ displays the smallest variance and bias overall in the identified transfer functions. This is also reflected in the standard deviation of the VAF values and is consistent with the conclusions in [38]. The results for the innovation estimation method CLMOESP show a slightly larger variance overall. This may be because of ill-conditioned projections and poor scaling related to the observability matrix of unstable systems [23]. Note, in this context, that systems 2–4 are indeed highly unstable with a pole far outside the unit disc; a situation that may not likely arise in practice. While the variances differ, all of the methods yield approximately unbiased estimates for this choice of $p = f = 4$. In Example 3, the innovation estimation method CLMOESP experiences some difficulties in terms of a large variance in the VAF on validation data.

It is interesting to note that in these four examples the OKID and direct parametrisation methods work quite reliably. This is in contrast to what will be observed in Section 5.3 where a high-order MIMO system is identified. It appears that differences between the identifications methods become far more pronounced when we are dealing with higher-order multivariable systems with an undefined system order. For low-order systems, these differences are minor.




doi: 10.1049/iet-cta.2012.0653

### 5.2 Numerical example: behaviour with poor excitation

In this example we consider identification of a simple fourth-order system with two inputs and two outputs, given by

$$A = \begin{bmatrix} 0.67 & 0.67 & 0 & 0 \\ -0.67 & 0.67 & 0 & 0 \\ 0 & 0 & -0.67 & -0.67 \\ 0 & 0 & 0.67 & -0.67 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.6598 & -0.5256 \\ 1.9698 & 0.4845 \\ 4.3171 & -0.4879 \\ -2.6436 & -0.3416 \end{bmatrix} \tag{39}$$

$$C = \begin{bmatrix} -0.3749 & 0.0751 & -0.5225 & 0.5830 \\ -0.8977 & 0.7543 & 0.1159 & 0.0982 \end{bmatrix}, \quad D = 0 \tag{40}$$

$$K = \begin{bmatrix} -0.6968 & -0.1474 \\ 0.1722 & 0.5646 \\ 0.6484 & -0.4660 \\ -0.9400 & 0.1032 \end{bmatrix} \tag{41}$$

To avoid the complication that the excitation of the input signal is affected by feedback, we consider an open-loop situation. The system is simulated from zero initial conditions for 1000 samples using the following input sequence

$$u_k = \begin{pmatrix} \sin\left(\frac{4}{10}\pi k\right) + \sin\left(\frac{11}{20}\pi k\right) \\ \sin\left(\frac{9}{20}\pi k\right) + \sin\left(\frac{6}{10}\pi k\right) \end{pmatrix} + v_k \tag{42}$$

$v_k$ is an additive Gaussian noise term with variance $8 \times 10^{-8}$. The innovation sequence $e_k$ is a Gaussian noise sequence with a variance of $10^{-4}$. Strictly speaking, the input signal is persistently exciting of any order, but it has been chosen such that the condition number of the VARX data matrix (cf. (9)) was of the order of $\varrho = 9 \times 10^3$. A past window $p = 5$ was chosen with a future window of $f = 5$ (e.g. the minimum dimensions to be able to estimate a fourth-order system with the CLMOESP method). Ten independent experiments were performed.

Fig. 4 shows a comparison of poles estimated using the CLMOESP and $\text{PBSID}_{\text{opt}}$ methods and Fig. 5 shows the identified transfer functions. While the poles seem to be identified consistently, the CLMOESP algorithm results in a smaller variance of the estimated poles. Also, the transfer functions are identified more accurately. This may have to to with a better numerical behaviour of the innovation estimation-type algorithm, in this case CLMOESP (cf. Section 2.2.4).

### 5.3 Experimental example: 'smart' beam dynamics

In this example, we consider the closed-loop identification of a 'smart' beam setup. This system is of interest since it is of a distributed-parameter nature and hence of potentially infinite order. Furthermore, within the considered bandwidth there are many resonances at frequencies an order of magnitude apart, such that this system could be considered to be a 'stiff' system.

Fig. 6 shows the setup that was used for experimental testing. The beam is approximately 1 m long and clamped at one end. It is equipped with six piezoelectric transducers (type M8528, from Smart Material Corp.), of which two are used for sensing, two for control and the two at the tip for introducing a disturbance which is to be rejected. The beam is controlled with an $\mathcal{H}_\infty$ controller that attempts to reject the disturbances injected at the tip.

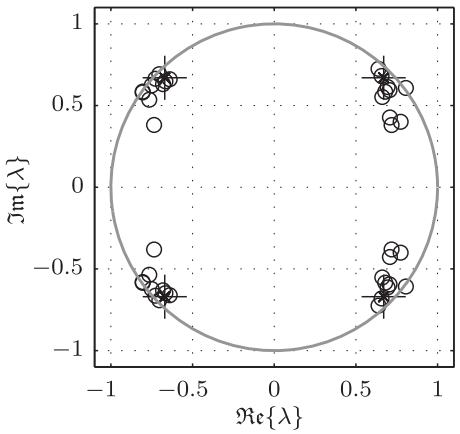


**Fig. 4** *Comparison of estimated poles using the CLMOESP (×) and $\text{PBSID}_{\text{opt}}$ (○) methods*

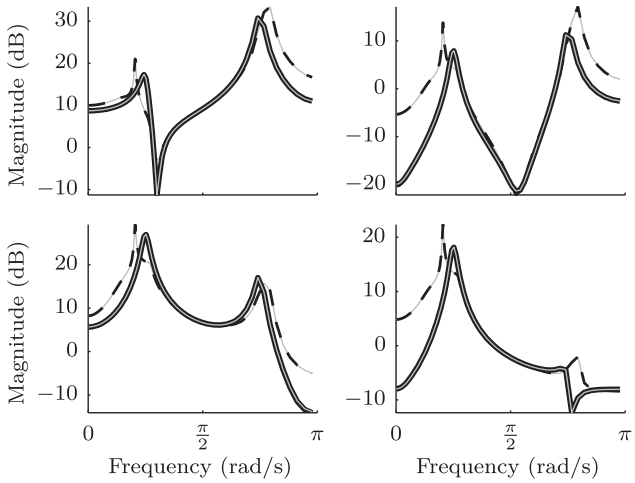


**Fig. 5** *Comparison of estimated models using the CLMOESP (solid grey) and $\text{PBSID}_{\text{opt}}$ (dashed) methods*

The true system is indicated with the black line

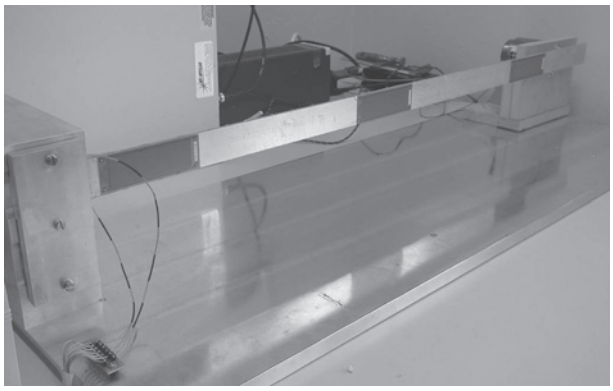

**Fig. 6** *'smart' beam setup*

A clamped beam is equipped with six piezoelectric transducers capable of either producing or sensing strain

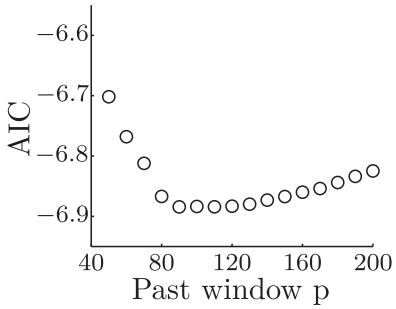


**Fig. 7** *AIC as a function of chosen past window size in the VARX step (Section 2.2.3)*

*5.3.1 Excitation and noise:* To be able to identify a model of the system in closed-loop, it is beneficial to inject a perturbation into the reference channels. [Although it may not be necessary, see e.g. [49], a higher signal to noise ratio results in a smaller variance of the identified model.] In principle we may choose either $\boldsymbol{r}_1$ or $\boldsymbol{r}_2$ or both to inject this perturbation (see Fig. 1). In the present example we inject a pseudorandom binary signal with an amplitude of 40 V into each of the two references. The switching probability was chosen so as to obtain a white perturbation for frequencies up to 200 Hz. Similar signals were generated to act as noise on the actuators at the tip. Here, an amplitude of 20 V was used. Signals were sampled at a rate of 1 kHz and $N = 4000$ and 16666 samples were used for identification and validation, respectively. In total, 65 of such experiments were performed, each with independent realisations of the perturbation and disturbance signals, allowing statistical properties of the estimates to be inferred. We remark that the data length for identification causes the data set to contain less than 15 cycles of the lowest natural frequency (3.7 Hz).

*5.3.2 Reference model:* As a reference model, we consider a non-parametric spectral estimate based on a long experiment ($N = 50\,000$ samples) performed under the same conditions as the other experiments. This estimate was subsequently averaged further over the 65 independent experiments so as to obtain an even smoother estimate. The estimate is obtained following the method in [97, 98]

$$\hat{\boldsymbol{G}}(e^{j\omega}) = \hat{\boldsymbol{\Phi}}_{yr}\hat{\boldsymbol{\Phi}}_{ur}^{-1} \tag{43}$$

where $\hat{\Phi}_{yr}$ is the estimated cross-spectrum between $\boldsymbol{y}_k$ and $\boldsymbol{r}_k$ and $\hat{\Phi}_{yr}$ is the estimated cross-spectrum between $\boldsymbol{u}_k$ and $\boldsymbol{r}_k$. This estimate can be shown to yield an asymptotically unbiased estimate provided the reference is persistently

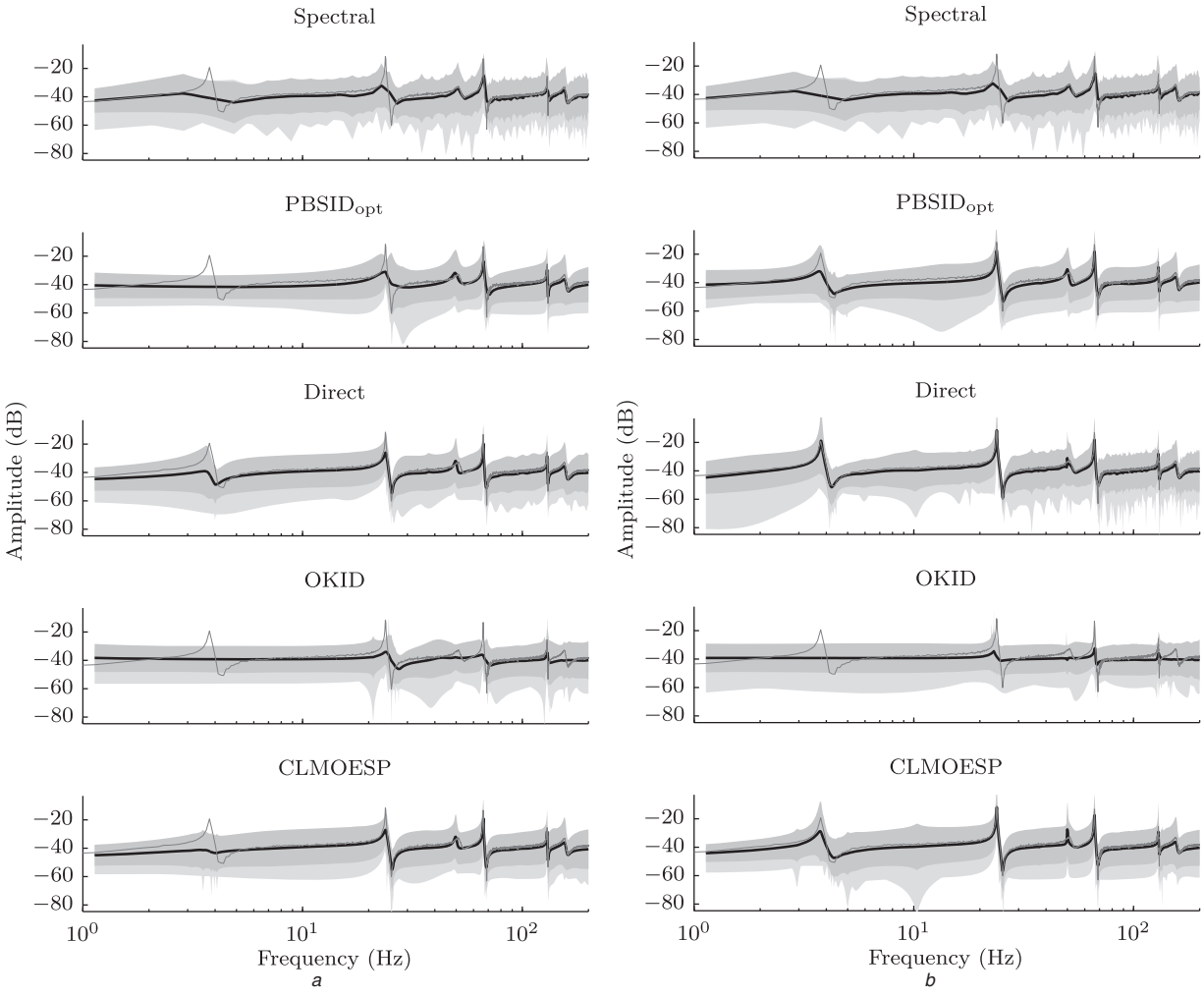


**Fig. 8** *Frequency response and confidence bounds; mean model (black solid), mean ±2σ (grey shaded), extreme models (light grey shaded) and reference model (grey solid)*

*a* $p = 60$
*b* $p = 200$



*IET Control Theory Appl.*, 2013, Vol. 7, Iss. 10, pp. 1339–1358
doi: 10.1049/iet-cta.2012.0653

exciting and uncorrelated with the noise signal. Obviously, the estimator will only give good estimates in the frequency range where $\boldsymbol{r}_k$ excites the system. We have computed this estimate using the recent 'local polynomial' method for non-parametric frequency response estimation [98]. This is a reliable method, which significantly reduces the adverse effects of spectral leakage. Typically, a comparison with a closed-loop spectral estimate is one powerful means of validating identified LTI models.

*5.3.3 Results:* The most critical tuning parameter common to the subspace algorithms is the choice of the past window size $p$ (equivalently, the VARX model order). Well-known tools in prediction-error identification can be used to choose the value of $p$ [1, 2, 23, 24, 78]. One such tool is the Akaike information criterion (AIC) [1]. For the VARX regression step we have shown the average AIC over ten experiments in Fig. 7 as a function of $p$. The AIC clearly suggests an order in the vicinity of $p = 100$.

To investigate the effect of the parameter $p$ and the validity of the value $p = 100$ suggested by the AIC, identification was performed for $p = \{60, 100, 200\}$. It is clear that the asymptotic variance of the VARX parameters will grow when the least-squares information matrix becomes ill-conditioned. This strongly depends on the amount of samples available and the richness of the signal $\boldsymbol{z}_k$, as discussed in Section 2.2.3.

A choice of $p \geq 60$ turned out to give a fairly 'white' innovation estimate. For simplicity, the future window $f$ in the $\text{PBSID}_{\text{opt}}$, OKID and CLMOESP methods was taken equal to $p$, but note that the choice of $f$ also affects the variance of the estimates, as discussed in Section 3.3 and in [46], in particular when the input spectrum exhibits zeros near the unit circle.

Using the order detection mechanisms (SVD) of the $\text{PBSID}_{\text{opt}}$ and CLMOESP methods a model order of $n = 29$ was found to give good prediction capability on a validation dataset. This order was also selected for the SVD truncation in the OKID method, since order detection in that method itself was rather unclear. Figs. 8*a* and *b* show the estimated confidence bounds resulting from the non-parametric estimate and the four key discrete-time methods ($\text{PBSID}_{\text{opt}}$, Direct parametrisation, OKID and CLMOESP) applied to 65 independent datasets for $p = \{60, 200\}$. Although the beam system is a two-input-two-output MIMO system, only the responses from input 1 to output 1 are shown.

Fig. 9 shows how well the estimated models predict the system's output on a validation dataset in terms of the VAF. The figures show that

(i) the prediction accuracy increases with past window $p$ for the $\text{PBSID}_{\text{opt}}$ method in particular;
(ii) the OKID method delivers unreliable estimates overall, where the other methods provide good models. A large choice of $p$ is required to obtain a reasonable model. This is related to the direct decomposition of the matrix in (14);
(iii) the VARX direct parameterisation method exhibits more outliers and a larger variance than the methods $\text{PBSID}_{\text{opt}}$ and CLMOESP, probably because of the lack of a model reduction step;
(iv) the spectral estimation method cannot capture the first resonance mode, probably due the the short duration (4 s) of the data records;
(v) a large value of $p$ helps to capture the first resonance mode. This may also be related to increasing the value of $f$ (note that we have chosen $f = p$).

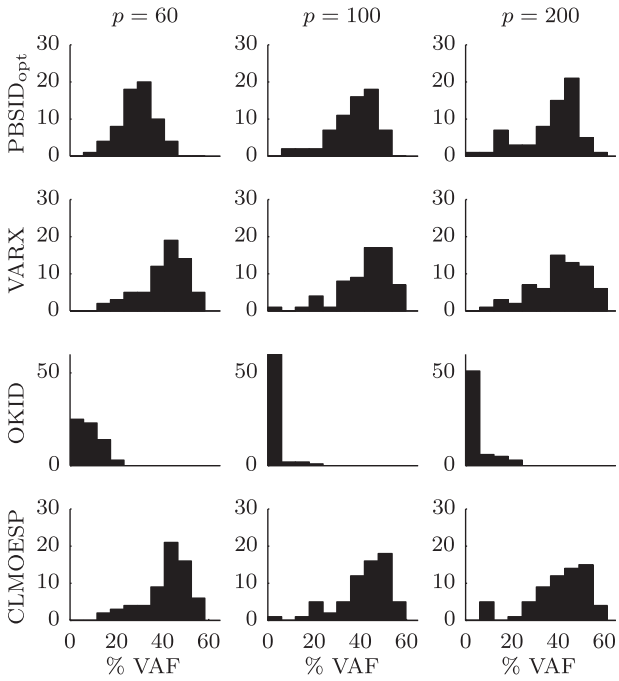


**Fig. 9** *Histograms of model quality in terms of output prediction: histogram of attained VAFs*

The results show that for $p = 100$ all resonances are captured and that the VAF improves when moving from $p = 60$ to 100 (much less when going from $p = 100$ to $p = 200$). This underlines the usefulness of the AIC as an order selection tool in the VARX modelling step. Although a choice of $p = 100$ was suggested by the AIC, it can be seen in Figs. 8*a* and *b* that the variance of the VARX model increases marginally for larger $p$; this trend was found to hold up to a past window of approx. $p = 700$, at which point the information matrix quickly became ill-conditioned. Of course, this highly depends on the character of $\boldsymbol{z}_k$ and the amount of samples.

### 5.4 Experimental example: continuous-time identification

Since the flexible beam setup of the previous example represents a stiff system with resonant frequencies an order of magnitude apart, the continuous-time formulation of $\text{PBSID}_{\text{opt}}$ was applied to the same dataset as in the previous section. The sample frequency (1 kHz) and number of samples ($N = 4000$) were kept equal to those of the discrete-time tests to allow for a proper comparison. For the continuous-time algorithm we chose $a = 240$ for the Laguerre filter time constant. Furthermore, a value of $p = 38$ was used for the past window size. Note that in the continuous-time case this value does not correspond to a time shift of $p$ samples, but to a projection operation that is iterated $p$ times.

Fig. 10 shows the estimated transfer function from input 1 to output 1 and its variance on the basis of 65 independent experiments (cf. Figs. 8*a* and *b*). While the figure shows that the variance is slightly higher than for the discrete-time $\text{PBSID}_{\text{opt}}$ algorithm, it also shows that the first resonance modes are captured very well, regardless of the fact that the experiment is short (4 s) and sampled at a rate (1 kHz), which is large in relation to the frequency of this mode (3.7 Hz). The discrete-time algorithm required a large value of $p$ and

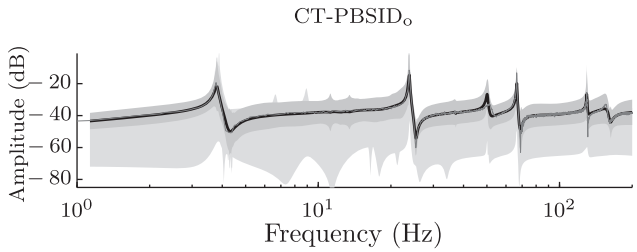


**Fig. 10** *Frequency response and confidence bounds; mean model (black solid), mean ±2σ (grey shaded), extreme models (light grey shaded) and reference model (grey solid)*

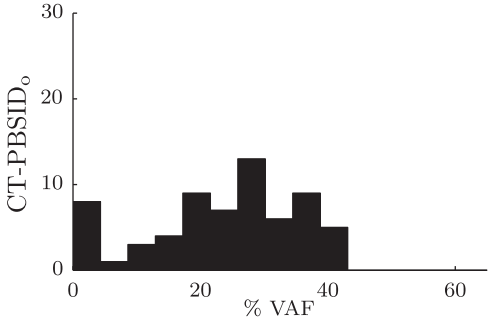


**Fig. 11** *Frequency response and confidence bounds; mean model (black solid), mean ±2σ (grey shaded), extreme models (light grey shaded) and reference model (grey solid)*

a longer dataset to properly capture these modes. On the other hand, the continuous time algorithm has more difficulty with the higher frequencies, which is reflected in the fact that there is a small bias at higher frequencies and not all the resonances are accurately captured. The corresponding histogram for the VAF is depicted in Fig. 11.

## 6 Concluding remarks

In this paper, an attempt has been made to organise the wide range of closed-loop subspace methods that has appeared over the last decade, both for the identification of discrete-time and continuous-time systems. Most of the algorithms can be derived from a few fundamental steps, which in turn can be traced back to autoregressive (VARX) modelling.

Based on experimental data obtained in repeated measurements several characteristics of the methods that are highly relevant in a practical context have been demonstrated. It turns out that the PBSID$_{opt}$ method [2] is a reliable method. The innovation estimation methods [3], among which CLMOESP [5], are of interest because of better numerical conditioning in the case of poorly exciting signals and their accurate order indication. The realisation-based approaches can only be used reliably if sufficiently large Hankel matrices can be constructed and may regularly incorrectly estimate the stability of resonant systems. Direct parametrisation of the VARX parameters has its own value in the sense that uncertainty on the parameters is readily characterised, but the models are not minimal and exhibit larger variance because of lack of a model reduction step.

The experimental results also demonstrated the value of continuous-time methods for their ability to deal with fast sampled data and stiff systems. On the other hand, these methods seem to suffer from a a slightly larger variance.

## 7 Acknowledgments

The work of, G.J. van der Veen was supported by Vestas Wind Systems A/S. The authors thank Ivo Houtzager for valuable discussions.

## 8 References

1 Ljung, L.: 'System identification: theory for the user' (Prentice-Hall PTR, 1999, 2 edn.)
2 Chiuso, A.: 'The role of vector autoregressive modeling in predictor-based subspace identification', *Automatica*, 2007, **43**, (6), pp. 1034–1048
3 Qin, S.J., Ljung, L.: 'Closed-loop subspace identification with innovation estimation'. IFAC Symp. on System Identification, Rotterdam, 2003
4 Di Ruscio, D.: 'Closed and open loop subspace system identification of the Kalman filter', *Model. Identif. Control*, 2009, **30**, (2), pp. 71–86
5 van der Veen, G.J., van Wingerden, J.W., Verhaegen, M.: 'Closed-loop MOESP subspace model identification with parametrisable disturbances'. Proc. 49th IEEE Control Conf. on Decision and Control, Atlanta, 12 2010
6 Gevers, M.: 'Identification for control: from the early achievements to the revival of experiment design', *Eur. J. Control*, 2005, **11**, (4–5), pp. 335–352
7 Verhaegen, M.: 'Application of a subspace model identification technique to identify LTI systems operating in closed-loop', *Automatica*, 1993, **29**, pp. 1027–1040
8 Chou, C.T., Verhaegen, M.: 'Subspace algorithms for the identification of multivariable dynamic errors-in-variables models', *Automatica*, 1997, **33**, (10), pp. 1857–1869
9 Van Overschee, P., De Moor, B.: 'Closed loop subspace system identification'. Proc. 36th IEEE Conf. on Decision and Control, 1997, vol 2, pp. 1848–1853
10 Ljung, L., McKelvey, T.: 'Subspace identification from closed loop data', *Signal Process.*, 1996, **52**, (2), pp. 209–215
11 Ng, T.S., Goodwin, G.C., Anderson, B.D.O.: 'Identifiability of mimo linear dynamic systems operating in closed loop', *Automatica*, 1977, **13**, (5), pp. 477–485
12 Van den Hof, P.M.J., de Callafon, R.A.: 'Multivariable closed-loop identification: from indirect identification to dual-youla parametrization'. Proc. 35th IEEE Conf. on Decision and Control, 1996, vol 2, pp. 1397–1402
13 Van den Hof, P.M.J.: 'Closed-loop issues in system identification', *Annu. Rev. Control*, 1998, **22** pp. 173–186
14 Jansson, M.: 'Subspace identification and ARX modeling'. Proc. 13th IFAC Symp. on System Identification, Rotterdam, Netherlands, 2003, pp. 1625–1630
15 Chiuso, A., Picci, G.: 'Constructing the state of random processes with feedback'. Proc. 13th IFAC Symp. on System Identification, Rotterdam, The Netherlands, 2003
16 Chiuso, A., Picci, G.: 'Geometry of oblique splitting subspaces, minimality and Hankel operators', in Rantzer, A., Byrnes, C. (Eds.): 'Directions in mathematical systems theory and optimization', vol 286 of (*Lecture Notes in Control and Information Sciences*), (Springer, Berlin, Heidelberg, 2003), pp. 85–126
17 Qin, S.J., Lin, W., Ljung, L.: 'A novel subspace identification approach with enforced causal models', *Automatica*, 2005, **41**, (12), pp. 2043–2053
18 Chiuso, A.: 'Asymptotic variance of closed-loop subspace identification methods', *IEEE Trans. Autom. Control*, 2006, **51**, (8), pp. 1299–1314,
19 Oku, H., Fujii, T.: 'Direct subspace model identification of LTI systems operating in closed-loop'. Proc. 43rd IEEE Conf. Decision and Control, vol. 2, 2004, pp. 2219–2224
20 Katayama, T., Kawauchi, H., Picci, G.: 'Subspace identification of closed loop systems by the orthogonal decomposition method', *Automatica*, 2005, **41**, (5), pp. 863–872
21 Gilson, M., Mercère, G.: 'Subspace based optimal IV method for closed loop system identification'. IFAC Symp. on System Identification, 2006
22 Gilson, M., Garnier, H., Young, P.C., Van den Hof, P.M.J.: 'Optimal instrumental variable method for closed-loop identification', *Control Theory Appl. IET*, 2011, **5**, (10), pp. 1147–1154
23 Chiuso, A., Picci, G.: 'Consistency analysis of some closed-loop subspace identification methods', *Automatica*, 2005, **41**, (3), pp. 377–391
24 Peternell, K., Scherrer, W., Deistler, M.: 'Statistical analysis of novel subspace identification methods', *Signal Process.*, 1996, **52**, (2), pp. 161–177




doi: 10.1049/iet-cta.2012.0653

25 Jansson, M., Wahlberg, B.: 'On consistency of subspace methods for system identification', *Automatica*, 1998, **34**, (12), pp. 1507–1519
26 Knudsen, T.: 'Consistency analysis of subspace identification methods based on a linear regression approach', *Automatica*, 2001, **37**, (1), pp. 81–89
27 Klein, V., Morelli, E.A.: 'Aircraft system identification: theory and practice' (AIAA, 2006)
28 Tischler, M., Remple, R.: 'Aircraft and rotorcraft system identification: engineering methods with flight-test examples' (AIAA, 2006)
29 Bergamasco, M., Lovera, M.: 'Continuous-time predictor-based subspace identification for helicopter dynamics'. '37th European Rotorcraft Forum', Gallarate, Italy, 2011
30 Garnier, H., Wang, L. (Eds.): 'Identification of continuous-time models from sampled data', (Springer, 2008)
31 Garnier, H., Soderstrom, T., Yuz, J.I.: 'Editorial special issue on continuous-time model identification', *Control Theory Appl. IET*, 2011, **5**, (7), pp. 839–841
32 Mohd-Moktar, R., Wang, L.: 'Continuous-time state space model identification using closed-loop data'. Second Asia Int. Conf. on Modelling & Simulation, Kuala Lumpur, Malaysia, 2008
33 Bergamasco, M., Lovera, M.: 'Continuous-time subspace identification in closed-loop'. Proc. 19th Int. Symp. on Mathematical Theory of Networks and Systems, Budapest, Hungary, 2010
34 Bergamasco, M., Lovera, M.: 'Continuous-time subspace identification in closed-loop using Laguerre filters'. Proc. 49th IEEE Control Conf. on Decision and Control Atlanta, 2010
35 Bergamasco, M., Lovera, M.: 'Continuous-time predictor-based subspace identification using Laguerre filters', *Control Theory Appl. IET*, 2011, **5**, (7), pp. 856–867
36 Haverkamp, L.R.J.: 'State space identification: theory and practice'. PhD thesis, Delft University of Technology, 2001
37 Ohta, Y., Kawai, T.: 'Continuous-time subspace system identification using generalized orthonormal basis functions'. Proc. 16th Int. Symp. on Mathematical Theory of Networks and Systems, Leuven, Belgium, 2004
38 Chiuso, A.: 'On the relation between CCA and predictor based subspace identification', *IEEE Trans. Autom. Control*, 2007, **52**, (10), 1795–1812
39 Enqvist, M.: 'Invertible time series and system identification in a nonlinear closed-loop setting', 'European research network on system identification (ERNSI Workshop)', Maastricht, Netherlands, 2012
40 Katayama, T.: 'Subspace methods for system identification' (Springer, 2005)
41 Van Overschee, P., De Moor, B.L.R.: 'Subspace identification for linear systems: theory, implementation, applications' (Kluwer Academic Publishers Group, Dordrecht, The Netherlands, 1996)
42 Goethals, I., Van Gestel, T., Suykens, J.A.K., Van Dooren, P., De Moor, B.: 'Identification of positive real models in subspace identification by using regularization', *IEEE Trans. Autom. Control*, 2003, **48**, (10), pp. 1843–1847, 10
43 Verhaegen, M., Verdult, V.: 'Filtering and system identification: a least squares approach' (Cambridge University Press, Cambridge, UK, 2007, 1st edn.)
44 Van den Hof, P.M.J., Schrama, R.J.P.: 'Identification and control – closed-loop issues', *Automatica*, 1995, **31**, (12), pp. 1751–1770
45 Di Ruscio, D.: 'Subspace system identification of the Kalman filter', *Model. Identif. Control*, 2003, **24**, (3), pp. 125–157
46 Chiuso, A.: 'On the asymptotic properties of closed-loop CCA-type subspace algorithms: Equivalence results and role of the future horizon', *IEEE Trans. Autom. Control*, 2010, **55**, (3), pp. 634–649
47 Katayama, T., Tanaka, H.: 'An approach to closed-loop subspace identification by orthogonal decomposition', *Automatica*, 2007, **43**, (9), pp. 1623–1630
48 Golub, G.H., Van Loan, C.F.: 'Matrix computations' (The Johns Hopkins University Press, 1996, 3rd edn.)
49 Bazanella, A.S., Gevers, M., Mišković, L.: 'Closed-loop identification of MIMO systems: a new look at identifiability and experiment design', *Eur. J. Control*, 2009, **16/3**, pp. 228–239
50 Gevers, M., Bazanella, A.S., Bombois, X.J.A., Mišković, L.: 'Identification and the information matrix: how to get just sufficiently rich? *IEEE Trans. Autom. Control*, 2009, **54**, (12), pp. 2828–2840
51 Bombois, X.J.A., Scorletti, G., Gevers, M., Van den Hof, P.M.J., Hildebrand, R.: 'Least costly identification experiment for control', *Automatica*, 2006, **42**, (10), pp. 1651–1662
52 Kulcsár, B., Verhaegen, M.: 'Robust cautious data driven control with guaranteed mean square stability'. Proc. 49th IEEE Control Conf. on Decision and Control, Atlanta, 2010, pp. 2276–2281
53 Dong, J., Verhaegen, M., Holweg, E.: 'Closed-loop subspace predictive control for fault tolerant MPC design'. Proc. 17th IFAC World Congress, vol. 17, 2008
54 Ho, B.L., Kalman, R.E.: 'Effective construction of linear state–variables models from input/output functions', *Regelungstechnik*, 1966, **14**, pp. 545–548
55 Kung, S.K.: 'A new low-order approximation algorithm via singular value decomposition'. Proc. 29th Asilomar Conf. on Circuits, Systems and Computers, 1978, pp. 705–714
56 Juang, J.N., Pappa, R.S.: 'An eigensystem realization algorithm (ERA) for modal parameter identification and model reduction', *AIAA, J. Guid. Control Dyn.*, 1985, **8**, (5), pp. 620–627
57 Phan, M., Horta, L.G., Juang, J.-N., Longman, R.W.: 'Linear system identification via an asymptotically stable observer', *J. Opt. Theory Appl.*, 1993, **79**, pp. 59–86, doi :10.1007/BF00941887
58 Phan, M., Horta, L.G., Juang, J.-N., Longman, R.W.: 'Improvement of observer/Kalman filter identification (OKID) by residual whitening', *J. Vib. Acoust.*, 1995, **117**, (2), pp. 232–239
59 Gustafsson, T.: 'Subspace-based system identification: weighting and pre-filtering of instruments', *Automatica*, 2002, **38**, (3), pp. 433–443
60 Bauer, D., Ljung, L.: 'Some facts about the choice of the weighting matrices in Larimore type of subspace algorithms', *Automatica*, 2002, **38**, (5), pp. 763–773
61 Di Ruscio, D.: 'A bootstrap subspace identification method: comparing methods for closed loop subspace identification by Monte Carlo simulations', *Model. Identif. Control*, 2009, **30**, (4), pp. 203–222
62 Chiuso, A., Muradore, R., Marchetti, E.: 'Dynamic calibration of adaptive optics systems: A system identification approach', *IEEE Trans. Control Syst. Technol.*, 2010, **8**, (3), pp. 705–713
63 Houtzager, I., van Wingerden, J.W., Verhaegen, M.: 'Fast-array recursive closed-loop subspace model identification'. Proc. 15th IFAC Symp. on System Identification, Saint Malo, France, 2009
64 Mercère, G., Lecœuche S., Lovera, M.: 'Recursive subspace identification based on instrumental variable unconstrained quadratic optimization', *Int. J. Adapt. Control Signal Process.*, 2004, **18**, pp. 771–797
65 Qin, S.J., Ljung, L.: 'Closed-loop subspace identification with innovation estimation'. Technical Report LiTH-ISY-R-2559, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, 12 2003
66 de Korte, R.B.C.: 'Subspace-based identification techniques for a 'smart' wind turbine rotor blade: a study towards adaptive data-driven control'. Master's thesis, Delft University of Technology, 1 2009
67 van der Veen, G.J., van Wingerden, J.W., Verhaegen, M.: 'Closed-loop MOESP subspace model identification with parametrisable disturbances'. Technical Report 10-015, Delft Center for Systems and Control, Delft University of Technology, 2010
68 Verhaegen, M.: 'Subspace model identification Part 3. Analysis of the ordinary output-error state-space model identification algorithm', *Int. J. Control*, 1993, **58**, (3), pp. 555–586
69 Brewer, J.W.: 'Kronecker products and matrix calculus in system theory', *IEEE Trans. Circuits Syst.*, 1978, **25**, (9), pp. 772–781
70 Verhaegen, M., Varga, A.: 'Some experience with the MOESP class of subspace model identification methods in identifying the BO105 helicopter'. Technical report, DLR Oberpfaffenhofen, 1994
71 Di Ruscio, D.: 'Subspace system identification of the Kalman filter'. Technical Report, Telemark University College, Porsgrunn, Norway, 2004
72 Nilsen, G.W.: 'Topics in open and closed loop subspace system identification: finite data based methods'. PhD thesis, Norwegian University of Science and Technology, 2005
73 Di Ruscio, D.: 'A method for the identification of state space models from input and output measurements', *Model. Identif. Control*, 1995, **16**, (3), pp. 129–143
74 Di Ruscio, D.: 'Combined deterministic and stochastic system identification and realization: DSR – a subspace approach based on observations', *Model. Identif. Control*, 1996, **17**, (3), pp. 193–230
75 Chiuso, A., Pillonetto, G., De Nicolao, G.: 'Subspace identification using predictor estimation via Gaussian regression'. Proc. 47th IEEE Conf. on Decision and Control, 2008, pp. 3299–3304, 12
76 Houtzager, I., van Wingerden, J.W., Verheagen, M.: 'Predictor-based subspace identification toolbox version 0.4', 2010
77 Houtzager, I., van Wingerden, J.W., Verhaegen, M.: 'VARMAX-based closed-loop subspace model identification'. Proc. 48th IEEE Conf. on Decision and Control, 2009, pp. 3370–3375
78 Kuersteiner, G.M.: 'Automatic inference for infinite order vector autoregressions', *Econ. Theory*, 2005, **21**, pp. 85–115
79 van Wingerden, J.W.: 'The asymptotic variance of the PBSIDopt algorithm'. Proc. 16th IFAC Symp. on System Identification, Brussels, 2012
80 Deistler, M., Peternell, K., Scherrer, W.: 'Consistency and relative efficiency of subspace methods', *Automatica*, 1995, **31**, (12), pp. 1865–1875

81 Bauer, D., Jansson, M.: 'Analysis of the asymptotic properties of the MOESP type of subspace algorithms', *Automatica*, 2000, **36**, (4), pp. 497–509

82 Van Gestel, T., Suykens, J.A.K., Van Dooren, P., De Moor, B.: 'Identification of stable models in subspace identification by using regularization', *IEEE Trans. Autom. Control*, 2001, **46**, (9), pp. 1416–1420

83 Miller, D.N., de Callafon, R.A.: 'Subspace identification with eigenvalue constraints', *Automatica*, 2013, in press

84 Lyzell, C., Enqvist, M., Ljung, L.: 'Handling certain structure information in subspace identification'. Proc. 15th IFAC Symp. on System Identification, 2009

85 Trnka, P., Havlena, V.: 'Subspace like identification incorporating prior information', *Automatica*, 2009, **45**, (4), pp. 1086–1091

86 Alenany, A., Shang, H., Soliman, M., Ziedan, I.: 'Brief paper – improved subspace identification with prior information using constrained least squares', *Control Theory Appl. IET*, 2011, **5**, (13), pp. 1568–1576

87 Liu, Z., Vandenberghe, L.: 'Interior-point method for nuclear norm approximation with application to system identification', *SIAM, J. Matrix Anal. Appl.*, 2009, **31**, (3), pp. 1235–1256

88 Hansson, A., Liu, Z., Vandenberghe, L.: 'Subspace system identification via weighted nuclear norm optimization', CoRR, abs/1207.0023, 2012

89 Bergamasco, M.: 'Continuous-time model identification with applications to rotorcraft dynamics'. PhD thesis, Politecnico di Milano, 2012

90 Bastogne, T., Garnier, H., Sibille, P.: 'A PMF-based subspace method for continuous-time model identification: application to a multivariable winding process', *Int. J. Control*, 2001, **74**, (2), pp. 118–132

91 Mercère, G., Ouvrard, R., Gilson, M., Garnier, H.: 'Identification de systèmes multivariables à temps continu par approche des sous-espaces', *J. Eur. Syst. Autom.*, 2008, **42**, (2–3), pp. 261–285

92 Johansson, R., Verhaegen, M., Chou, C.T.: 'Stochastic theory of continuous-time state-space identification', *Signal Process. IEEE Trans.*, 1999, **47**, (1), pp. 41–51

93 Ohta, Y., Kawai, T.: 'Continuous-time subspace system identification using generalized orthonormal basis functions'. 16th Int. Symp. on Mathematical Theory of Networks and Systems, Leuven, Belgium, 2004

94 Ohta, Y.: 'Realization of input–output maps using generalized orthonormal basis functions', *Syst. Control Lett.*, 2005, **54**, (6), pp. 521–528

95 Kinoshita, Y., Ohta, Y.: 'Continuous-time system identification using compactly supported filter kernels generated from Laguerre basis functions'. Proc. 49th IEEE Control Conf. on Decision and Control, Atlanta, 2010

96 Bergamasco, M., Lovera, M., Ohta, Y.: 'Recursive continuous-time subspace identification using Laguerre filters'. 50th IEEE Conf. on Decision and Control and European Control Conference, Orlando, USA, 2011

97 Akaike, H.: 'Some problems in the application of the cross-spectral method', in Harris, B. (Ed.): 'Spectral analysis of time series' (Wiley, 1967), pp. 81–107

98 Pintelon, R., Schoukens, J., Vandersteen, G., Barbé, K.: 'Estimation of nonparametric noise and FRF models for multivariable systems – Part I: Theory', *Mech. Syst. Signal Process.*, 2010, **24**, (3), pp. 573–595





doi: 10.1049/iet-cta.2012.0653